(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 June 2001 (21.06.2001)

PCT

(10) International Publication Number
WO 01/44512 A2

(51) International Patent Classification[7]: C12Q 1/68, G01N 33/50, C07K 14/47, C12N 9/00, 15/63, 5/10, C07K 16/18, A61K 31/7088

(21) International Application Number: PCT/US00/34027

(22) International Filing Date:
15 December 2000 (15.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/171,034	16 December 1999 (16.12.1999)	US
Not furnished	6 December 2000 (06.12.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/171,034 (CIP)
Filed on	16 December 1999 (16.12.1999)
US	Not furnished (CIP)
Filed on	6 December 2000 (06.12.2000)

(71) Applicant *(for all designated States except US)*: **CURAGEN CORPORATION** [US/US]; 555 Long Wharf Drive, 11th floor, New Haven, CT 06511 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **GOULD-ROTHBERG, Bonnie, E.** [CA/US]; 1701 Moose Hill Road, Guilford, CT 06437 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky, and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/44512 A2

(54) Title: METHOD OF IDENTIFYING LIGANDS FOR THE PEROXISOME PROLIFERATOR ACTIVATED RECEPTOR GAMMA USING DIFFERENTIAL GENE EXPRESSION

(57) Abstract: Disclosed are methods of identifying ligands for the peroxisome proliferator activated receptor gamma (PPARγ) using differential gene expression. Also disclosed are novel nucleic acid sequences whose expression is differentially regulated by PPARγ ligands.

METHOD OF IDENTIFYING LIGANDS FOR THE PEROXISOME PROLIFERATOR ACTIVATED RECEPTOR GAMMA USING DIFFERENTIAL GENE EXPRESSION

FIELD OF THE INVENTION

The invention relates generally to nucleic acids and polypeptides and in particular to the identification of ligands for the Peroxisome Proliferator Activated Receptor Gamma (PPARγ) using differential gene expression.

BACKGROUND OF THE INVENTION

Diabetes is known to affect approximately eight million people in the United States. Over 90% of diagnosed diabetics suffer from noninsulin-dependent diabetes mellitus (NIDDM), which is also known as type II diabetes. An additional eight million people may have undiagnosed NIDDM. Obesity, advancing age, a family history of NIDDM, a sedentary lifestyle, a history of gestational diabetes, and the presence of co-morbid conditions such as hypertension or hyperlipidemia are risk factors for this disorder.

NIDDM is associated with functional and biochemical abnormalities in the pancreas, liver and peripheral insulin-sensitive tissues such as skeletal muscle and adipose tissue. The abnormalities can include, *e.g.* relative, but not absolute deficiency of pancreatic insulin secretion, an increased rate of hepatic glucose production and extreme insulin resistance in peripheral tissues such as adipose and skeletal muscle.

One hypothesis for the pathogenesis of NIDDM suggests that the initial event is not pancreatic failure but the development of peripheral tissue insulin resistance. During this "pre-diabetic state" candidate NIDDM patients actually demonstrate hyperinsulinemia, which is an increase level of insulin in the plasma due to an increase in secretion of insulin by the beta cells of the pancreatic islets. The observed pancreatic insulin deficiency that follows is most likely related to pancreatic burnout from maintaining the hyperinsulinemic state.

Persistent, untreated hyperglycemia can result in, *e.g.*, increased risk of urinary tract infections and dehydration related to polyuria. However, often the most important sequellae of diabetes are its long term complications. Following 15-20 years of poorly managed diabetes,

patients are at risk for peripheral vascular disease with risk of limb-amputating gangrene, blindness, myocardial infarction and renal failure.

One class of therapeutics used to control of NIDDM is the thiazolidenedione compounds. These compounds have been classified as synthetic ligands for the Peroxisome Proliferator Activated Receptor γ(PPARγ). PPARγ is a nuclear hormone receptor and it has been shown to have metabolic activity primarily in the peripheral tissues where insulin resistance takes place.

PPARγ receptors exist as three splice variants, γ1 and γ2 and: γ3. PPARγ1 has a ubiquitous tissue distribution with increased expression levels in the heart, liver and kidney. Its function in these tissues is largely uncharacterized. A natural ligand for PPARγ is 15-deoxy-Δ-J2 prostaglandin.

SUMMARY OF THE INVENTION

The invention is based in part on the discovery that certain nucleic acids are differentially expressed in liver tissue of animals treated with,N-(2-benzoylphenyl)-L-tyrosine, a synthetic Peroxisome Proliferator Activated Receptor Gamma ligand (PPARγL). These differentially expressed nucleic acids include novel sequences and nucleic acids sequences that, while previously described, have not heretofore been identified as PPARγ responsive.

In various aspects, the invention includes methods of identifying PPARγ ligands, methods of diagnosing PPARγ pathophysiologies, and methods of treating those pathophysiologies. For example, in one aspect, the invention provides a method of identifying a PPARγ ligand by providing a test cell population comprising a cell capable of expressing one or more nucleic acids sequences responsive to PPARγ ligands, contacting the test cell population with the test agent and comparing the expression of the nucleic acids sequences in the test cell population to the expression of the nucleic acids sequences in a reference cell population. An alteration in expression of the nucleic acids sequences in the test cell population compared to the expression of the gene in the reference cell population indicates that the test agent is a ligand for PPARγ.

The invention in a further aspect includes a method of selecting an individualized therapeutic agent appropriate for a particular subject. The method includes providing from the subject a test cell population comprising a cell capable of expressing one or more nucleic acids

sequences responsive to PPARγ ligands, contacting the test cell population with the therapeutic agent, and comparing the expression of the nucleic acids sequences in the test cell population to the expression of the nucleic acids sequences in a reference cell population. An alteration in expression of the nucleic acids sequences in the test cell population compared to the expression of the nucleic acids sequences in the reference cell population indicates the therapeutic agent is appropriate for the particular subject.

In a further aspect, the invention provides a method of diagnosing or determining susceptibility to a PPARγ mediated pathophysiology, *e.g.*, noninsulin-dependent diabetes mellitus, or adipocyte differentiation, liver failure, jaundice, or NH_3 toxicity. The method includes providing from the subject a cell population comprising a cell capable of expressing one or more PPARγ-responsive genes, and comparing the expression of the nucleic acids sequences to the expression of the nucleic acids sequences in a reference cell population that includes cells from a subject not suffering from a PPARγ mediated pathophysiology. An alteration in expression of the nucleic acid sequences in the test cell population compared to the expression of the nucleic acids sequences in the reference cell population indicates subject has or is susceptible to a PPARγ mediated pathophysiology.

Also provided are novel nucleic acids, whose expression is responsive to the effects of N-(2-benzoylphenyl)-L-tyrosine, as well as single nucleotide polymorphisms in HEPATO sequences, as well as a method of using the HEPATO single nucleotide polymorphisms.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is based in part on the discovery of changes in expression patterns of multiple nucleic acid sequences in rodent liver cells following exposure to a ligand for the Peroxisome Proliferator Activated Receptor Gamma (PPARγ). The differences in gene expression were identified following administration of 5.0 mg/kg/day of a PPARγ-specific ligand, N- (2-benzoylphenyl)-L-tyrosine, referred to herein as PPARγL. This compound is described in Henke et al., J. Med. Chem.41:5020-5036, 1998, whose contents are incorporated herein in their entirety. The nucleic acid sequences and polypeptides of the present invention are herein referred to as HEPATO.

The differentially expressed nucleic acids were identified by administering PPARγL to male 10-14 week old Sprague Dawley rats 5.0 mg/kg/day b.i.d. dosing for 3 days. Control animals received N-methylglucamine. The animals were sacrificed 24 hours following the last dose. Liver tissue was dissected from the animals, and total RNA was recovered from the dissected tissue. cDNA was prepared and the resulting samples were processed through using GENECALLING™ differential expression analysis as described in U. S. Patent No. 5,871,697 and in Shimkets et al., Nature Biotechnology 17:798-803 (1999). The contents of these patents and publications are incorporated herein by reference in their entirety.

Thirty-five gene fragments were found to be differentially expressed in rat liver tissue in response PPARγL. The sequences are referred to herein as HEPATO 1-35. A summary of the sequences analyzed is presented in Table 1.

Nine sequences (HEPATO: 1-9) represent novel rat genes for which the sequence identity to sequences found in public databases is either high (i.e., > 90%, observed for 3 fragments), moderate (i.e., between about 70% and about 90%, observed for 4 genes) or low (i.e., < 70%, observed for 1 fragment).

For nine of the nucleic acids (HEPATO: 10-18), no homology was found to known nucleic acid sequences in public databases.

16 of the nucleic acids identified have been previously described. 12 of the sequences (HEPATO19-30) are newly shown to be PPARγ responsive. Two sequences (HEPATO: 33-34) has been previously recognized as being differentially expresses as part of the PPARγ

response in adipose tissue. Two other sequences have been shown to be differentially expressed as part of the PPARα response in liver tissue (HEPATO: 31-36).

One sequence, HEPATO35, a novel sequence with 83% homolgy to murine glycerol-3-phosphate acyltransferase, has also been shown to be differentially expressed in adipose tissue in response to N- (2-benzoylphenyl)-L-tyrosine.

In addition, nine LINE elements have also been shown to be differentially expressed in hepatic tissue in response to N- (2-benzoylphenyl)-L-tyrosine. Eight of the LINE elements have been previously described (HEPATO: 36-43), and one is newly described (HEPATO44).

For some of the novel sequences (i.e., HEPATO: 10-18), a cloned sequence is provided along with one or more additional sequence fragments (*e.g.*, ESTs or contigs) which contain sequences substantially identical to, the cloned sequence. Also provided is a consensus sequence which includes a composite sequence assembled from the cloned and additional fragments. For a given HEPATO sequence, its expression can be measured using any of the associated nucleic acid sequences may be used in the methods described herein. For previously described sequences (HEPATO:20-35) database accession numbers are provided for both the rat sequences and their human homologues. The accession number allow one of ordinary skill in the art to deduce information necessary for detecting and measuring expression of the HEPATO nucleic acid sequences in a desired cell sample.

The PPARγ -responsive nucleic acids discussed herein include the following:

Table 1

Description of Sequence	Sequence Database Reference	PPARγL Effect onTranscript Level	HEPATO Assignment	SEQ ID NO
PPARγ Responsive Novel Nucleic Acid Sequences				
Novel gene fragment, 695 bp, 83% SI and 92% amino acid identity to cow B22 subunit of the NADH-ubiquinone oxidoreductase complex [X64836]	N/A	-2	HEPATO1	1-10
Novel gene fragment, 156 bp, 92% SI to mouse protein kinase inhibitor gamma [U97170]	N/A	+5	HEPATO2	11
Novel gene fragment, 1084 bp, 100% amino acid identity to rat hepatocarcinogenesis-related transcription factor [JC4857]	N/A	+2	HEPATO3	12-17
Novel gene fragment, 514 bp, 83% SI to human protein disulfide isomerase related protein [D49490]	N/A	+2	HEPATO4	18-22
Novel gene fragment, 165 bp, 67% SI to human cathepsin O [X77383]	N/A	+2.5	HEPATO5	23

Novel gene fragment, 136 bp, 83% SI to human complement protein C8 beta subunit [M16973]	N/A	+1.5	HEPATO6	24-26
Novel gene fragment, 789 bp, 88% SI to mouse complement factor I (C3b/C4b inactivator) [U47810]	N/A	+3	HEPATO7	27-35
Novel gene fragment, 141 bp, 89% SI to human KIAA0315 [AB002313]	N/A	+6	HEPATO8	36-38
Novel gene fragment, 270 bp, 64% amino acid similarity to rat HES-1 (hairy and enhancer of split) transcription factor [Q04666]	N/A	+4	HEPATO9	39-41
Novel gene fragment, 189 bp	N/A	+2	HEPATO10	42
Novel gene fragment, 584 bp	N/A	-100	HEPATO11	43-48
Novel gene fragment, 45 bp	N/A	-1.5	HEPATO12	49
Novel gene fragment, 63 bp	N/A	+5	HEPATO13	50
Novel gene fragment, 96 bp	N/A	+20	HEPATO14	51
Novel gene fragment, 315 bp	N/A	+2	HEPATO15	52-55
Novel gene fragment, 169 bp	N/A	+10	HEPATO16	56-58
Novel gene fragment, 391 bp	N/A	+2	HEPATO17	59-62
Novel gene fragment, 45 bp	N/A	-2	HEPATO18	63
Previously Described Nucleic Acid Sequences Newly Shown To Be PPAR γ Responsive in Liver Tissue				
Glyceraldehyde-3-phosphate dehydrogenase	M17701	+10	HEPATO19	
UDP-glucose dehydrogenase	AB013732	+3	HEPATO20	
Succinyl-CoA synthetase alpha subunit	J03621	-3	HEPATO21	
UDP-glucosuronosyl transferase	Y00156	+1.5	HEPATO22	
Alanine aminotransferase	D10354	-2	HEPATO23	
Tyrosine aminotransferase	M18340	-2	HEPATO24	
Catechol-O-methyltransferase	M60754	+4	HEPATO25	
Vitamin D binding protein	J05148	+1.5	HEPATO26	
Submaxillary gland alpha-2-mu globulin	J00738	+25	HEPATO27	
Phospholipase C-1	M20636	-2	HEPATO28	
Ribosomal protein L18	M20156	-2	HEPATO29	
Endoplasmic reticulum transmembrane protein	Y07783	+3	HEPATO30	
Cytochrome P450 IVA2	M57119	+6	HEPATO31	
Cytochrome p450 IVA3	M33936	+5	HEPATO32	
Nucleic Acid Sequences Previously Demonstrated To Be PPARγ Responsive in Adipose Tissue				
Tricarboxylate transport protein	L12016	+9	HEPATO33	
Cytochrome C oxidase polypeptide I-mitochondrial	S79304	+40	HEPATO34	
Gene fragment, 744 bp, 83% SI to mouse glycerol-3-phosphate acyltransferase [M77003]		-2	HEPATO35	

Previously Described Line Elements Newly Shown To Be PPAR γ Responsive in Liver Tissue				
L1 retrotransposon m1vi2-m38 [Found in CYP4A2 5' region 2x]	U87604	-15	HEPATO36	
L1 retrotransposon, ORF 2 [Found in CYP2A2]	X61295	-10	HEPATO37	
L1 RnB7 repetitive DNA element [Found in CYP4A2 5' region 2x]	X07687	-5	HEPATO38	
L1 retrotransposon m1vi2-m8 [Found in CYP4A2 5' region]	U87598	-8	HEPATO39	
LINE DNA containing 7 ORF's	X53581	-6	HEPATO40	
Long interspersed repetitive DNA element LINE3	M13100	-2	HEPATO41	
2.4 kb repeat DNA left terminal region	X50473	-100	HEPATO42	
Satellite I core DNA	M30113	-20	HEPATO43	
PPARγ Responsive Novel Line Element				
Novel LINE element, 326 bp, [Found in CYP2C13]	N/A	-100	HEPATO44	64-68

SI=Similar to

Below follows additional discussion of nucleic acid sequences whose expression is differentially regulated in the presence of PPARγL.

HEPATO1

HEPATO1 is a novel 695 bp gene fragment. The nucleic acid was initially identified in a poisoned fragment having the following sequence:

AGCTGCTGAGGCAGGCGGAAGAAGAATTCTGGCAAAACCAGCATCCTCAGCCGTATATCTTCCCAGACTCTCCTGGGGGT
ACTTCCTATGAGAGATACGAGTGCTACAAGGTTCCAGAATGGTGCTTAGATTACTGGCATCCTTCTGAGAAAGCAGTGTA
TCCTGATTACTTTTCCAAGAGAGAGCAGTGGAAGAAACTGAGGATGGAGAGCTGGGATCGGGAGGTTAAACAGCTGGAGG
AAGAAACGTCANCTGATGGTATTATGACTGAAGCTTTGCCTCCTGCCAGAAAGGAAGGCGACTTGCCCCCATTGTGGTGG
CATATTTTGACCAGACCTCGGGNACGGCCCACATAGGGACAGGCACCG (SEQ ID NO: 1).

The sequence was assembled into a contig that includes:

```
 1  agggcaagtg tgaatgtgtt ctgtaactct cacttgcaag catgaacagt gcggtgcctg
 61  tctctatgtg ggccgttccc gaggtctggt cacaatatgc caccacaatg ggggcaagtc
121  gccttccttt ctggcaggag gcaaagcttc agtcataata ccatcaggtg acgtttcttc
181  ctccagctgt ttaacctccc gatcccagct ctccatcctc agtttcttcc actgctctct
241  cttggaaaag taatcaggat acactgcttt ctcagaagga tgccagtaat ctaagcacca
301  ttctggaacc ttgtagcact cgtatctctc ataggaagta cccccaggag agtctgggaa
361  gatatacggc tgaggatgct ggttttgcca gaattcttct tccgcctgcc tca
(SEQ ID NO: 2);
```

```
  1  gtctctagtt acttttatta gggcaagtgt gaatgtgttc tgtaactctc acttgcaagc
 61  atgaacagtg cggtgcctgt ctctatgtgg gccgttcccg aggtctggtc acaatatgcc
121  accacaatgg gggcaagtcg ccttcctttc tggcaggagg caaagcttca gtcataatac
181  catcaggtga cgtttcttcc tccagctgtt taaccttccg atcccagctc tccatcctca
241  gtttcttcca ctgctctctc ttggaaaagt aatcaggata cactgctttc tcagaaggat
301  gccagtaatc taagcaccat tctggaacct tgtagcactc gtatctctca taggaagtac
361  ccccaggaga gtctgggaag atatacggct gaggatgctg gttttgccag aattcttctt
(SEQ ID NO: 3);

  1  ccagaatggt gcttagatta ctggcatcct tctgagaaag cagtgtatcc tgattacttt
 61  tccaagagag agcagtggaa gaaactgagg atggagagct gggatcggga ggttaaacag
121  ctggaggaag aaacgtcacc tgatggtatt atgactgang ctttgcctcc tgccagaaag
181  gaaggcgact tgcccccatt gtggtggcat attgtgacca gacctcggga acggcccaca
241  tagagacagg caccgcactg ttcatgcttg caagtgagag ttacagaaca cattcacact
301  tgccctaata aaagtaacta gagac (SEQ ID NO:4);

  1  ctagttactt ttattagggc aagtgtgaat gtgttctgta actctcactt gcaagcatga
 61  acagtgcggt gcctgtctct atgtgggccg ttcccgaggt ctggtcacaa tatgccacca
121  caatgggggc aagtcgcctt cctttctggc aggaggcaaa gcttcagtca taataccatc
181  aggtgacgtt tcttcctcca gctgtttaac ctcccgatcc cagctccaca tcctcagttt
241  cttccactgg tcc (SEQ ID NO: 5);

aagcttcagtcataataccatcaggtgacgtttcttcctccagctgtttaacctcccgat
cccagctctccatcctcagtttcttccactgctctctcttggaaaagtaatcaggataca
ctgctttctcagaaggatgccagtaatctaagcaccattctggaaccttgtagcactcgt
atctctcataggaagtacccccaggagagtctgggaagatatacggctgaggatgctggt
tttgccagaattcttcttccgcctgcctcagcagctgggtggccttcatcatgtccttct
cattcttatgttcttcaaaccgggctctcatcaagcaagcaaggtaccggtatttgtccc
tgtggacacaccatgactcgag (SEQ ID NO: 6);

nctatggcgttctgtnctcccccgncctacttaacccaccggcagaagntgctgcggctg
tataagcgcgcgctgccacctcgagtcatggtgtgtccacagggacaaataccggtacct
tgcttgcttgatgagagcccggtttgaagaacataagaatgagaaggacatgatgaaggc
cacccagctgctgaggcaggcggaagaagaattc (SEQ ID NO: 7);

gaattcttcttccgcctgcctcagcagctgggtggccttcatcatttttnttctcattct
tatgttctgtgctcccccggcctacttaacccaccgggcagaagggtgctgcgggctggt
attaaagccgccgccgcctgcgccacctcgagtcatggngtgtccacagggacaaaatac
ccggtacc (SEQ ID NO: 8);
```

and

```
gaattcttcttccgcctgcctcagcagctgggtggccttcatcatgtccttctcattctt
atgttcttcaaaccgggctctcatcaagcaagcaaggtaccgggtatttggtccctgtgg
gacacaccatgg (SEQ ID NO: 9).
```

The resulting consensus sequence is:

```
  1 GTCTCTAGTTACTTTTATTAGGGCAAGTGTGAATGTGTTCTGTAACTCTCACTTGCAAGCATGAACAGTGCGGTGCCTGT
 81 CTCTATGTGGGCCGTTCCCGAGGTCTGGTCACAATATGCCACCACAATGGGGGCAAGTCGCCTTCCTTTCTGGCAGGAGG
161 CAAAGCTTCAGTCATAATACCATCAGGTGACGTTTCTTCCTCCAGCTGTTTAACCTCCCGATCCCAGCTCTCCATCCTCA
241 GTTTCTTCCACTGCTCTCTCTTGGAAAAGTAATCAGGATACACTGCTTTCTCAGAAGGATGCCAGTAATCTAAGCACCAT
321 TCTGGAACCTTGTAGCACTCGTATCTCTCATAGGAAGTACCCCCAGGAGAGTCTGGGAAGATATACGGCTGAGGATGCTG
401 GTTTTGCCAGAATTCTTCTTCCGCCTGCCTCAGCAGCTGGGTGGCCTTCATCATGTCCTTCTCATTCTTATGTTCTTCAA
481 ACCGGGCTCTCATCAAGCAAGCAAGGTACCGGGTATTTGTCCCTGTGGACACACCATGACTCGAGGTGGCGCAGGCGGCG
561 GCGGCTTTAATACCAGCCCGCAGCACCCTTCTGCCCGGTGGGTTAAGTAGGCCGGGGGAGCACAGAACGCCATAAGNAAT
641 GAGAANAAAAATGATGAAGGCCACCCAGCTGCTGAGGCAGGCGGAAGAAGAATTC (SEQ ID NO: 10).
```

Its expression is decreased 2-fold in PPARγL treated rats.

HEPATO2

HEPATO2 is a novel 156 bp gene fragment.

The cloned sequence is:

```
  1 agatctctct cagcctgtga ggaccacagg cagactcagg atgaggtgtt cgcgtcactg
 61 ctctccggct ggctgctggc ttccttgtct ggggtacttc cctctgcctg tccttctgcc
121 ccctcaagtg cgagctcgcc catgtctcca gctagc (SEQ ID NO: 11).
```

Its expression is increased 5-fold in PPARγL treated rats.

HEPATO3 a novel 1084 bp, gene fragment

HEPATO3 is a novel 1084 bp gene fragment.

The nucleic acid was initially identified in a cloned fragment having the following sequence:

```
  1 tgatcaggga aggagcccag tggtgggtgg ctttagacac taatcagctg ggggaaaagt
 61 tcgttggcaa aagtgtcctc ccaggagtgg tctgtaccaa gtggagaaga catgtcgctg
121 aagggagaag gggagccctc atatccacag tcactgtgag catccagcag gcaggaaggt
181 ggtctcagac aatggctgga tgaaagcagg tttgagatgc cccgctntaa ccatgaagtc
241 ttcatccaaa ggttctttct tcctgagaca atgaattc (SEQ ID NO: 12).
```

The cloned sequence was assembled into a contig that includes:

```
  1  gaggtttgaa atactttaga ggaattctta taaggccaga tagcctctac tttggctttt
 61  agtgaagcag acaggcctat gctatcctct aggcagtgtg atgatcaggg aaggagccca
121  gtggtgggtg gctttagaca ctaatcagct gggggaaaag ttcgttggca aaagtgtcct
181  cccaggagtg gtctgtacca agtggagaag acatgtcgct gaagggagaa ggggagccct
241  catatccaca gtcactgtga gcatccagca ggcaggaagg tggtctcaga caatggctgg
```

```
301  atgaaagcag gtttgagatg cccagctcgg gaatgaagtc atcatccaaa ggttctttct
361  tcactgagac aatgaattca gggtgatcct cttctgaaga gcttagaggt gcttcctcaa
421  ttttcactac cacattagtt tggctctctg tctcagaggg gatctctaag actaaaggct
481  tggtgtatac atggtcaaaa cgaatgagtt ca (SEQ ID NO: 13);
```

```
  1  ggaattctta taaggccaga tagcctctac tttggctttt agtgaagcag acaggcctat
 61  gctatcctct aggcagtgtg atgatcaggg aaggagccca gtggtgggtg gctttagaca
121  ctaatcagct gggggaaaag ttcgttggca aaagtgtcct cccaggagtg gtctgtacca
181  agtggagaag acatgtcgct gaagggagaa ggggagccct catatccaca gtcactgtga
241  gcatccagca ggcaggaagg tggtctcaga caatggctgg atgaaagcag gtttgagatg
301  cccagctcgg gaatgaagtc atcatccaaa ggttctttct tcactgagac aatgaattca
361  gggtgatcct cttctgaaga gcttataggt gcttcctcaa ttttcactac cacattagtt
421  tggctctctg tctcagaggg gatctctaag actagaggct tggtgtatac atggtcaaaa
481  cgaatgagtt cattaatggc tt (SEQ ID NO: 14);
```

```
  1  cggccgcaaa ggtttttcaa atgtccttcc ccagagtctg ctaatctgga ggaactccca
 61  gaggtctacc cagaaggacc tagttcctta ccagcctccc tttctctgtc agtggggacc
121  tcatcagcca agctggaagc cattaatgaa ctcattcgtt ttgaccatgt atacaccaag
181  cctctagtct tagagatccc ctctgagaca gagagccaaa ctaatgtggt agtgaaaatt
241  gaggaagcac ctctaagctc ttcagaagag gatcaccctg aattcattgt ctcagtgaag
301  aaagaacctt tggatgatga cttcattccc gagctgggca tctcaaacct gctttcatcc
361  agccattgtc tgagaccacc ttcctgcctg ctggatgctc acagtgactg tggatatgag
421  ggctcccctt ctcccttcag cgacatgtct tctccacttg gtacagacca ctcctgggag
481  gacacttttg ccaacgaact tttccccca (SEQ ID NO: 15);
```

and

```
  1  cacagtaaat ttccaataat tttacaaaga ttcttgatct tcacttgaac tggacataag
 61  gaaggacagg cccctcaggt tgctgtttct ctgcttgtag aaggaaacaa aagaaacctg
121  tggggcgggg aggagagaaa gaactggtga ctctcatgtc tacttcagga catgtgaaga
181  ggccggtgtg gagctgcaca cctggtaaag tccagcactt gggagtgggg tcaagagggt
241  cacaagtttc agcttagcct cggctacata gccaggctga acgataactg tcagatgact
301  ttccctatga tttagagcat gctaccacct ttaagataat gagaatctca aaagctgtag
361  tattggaata cctttgaaga cctcagacag ctgagtgtca aaagacaata cttggaagtc
421  atctaagagg tttgaaatac tttagaggaa ttcttataag gccagatagc ctctactttg
481  gcttttagtg aagcagacag gcctatgcta tcctctaggc agtgtgatga tcagggaagg
541  agcccagtgg tgggtggctt tagacactaa tcagctgggg gaaaagttcg ttggcaaaag
601  tgtcctccca agagtggtct gtaccaagtg gagaagacat (SEQ ID NO: 16).
```

The resulting consensus sequence is:

```
   1 CGGCCGCAAAGGTTTTTCAAATGTCCTTCCCCAGAGTCTGCTAATCTGGAGGAACTCCCAGAGGTCTACCCAGAAGGACC
  81 TAGTTCCTTACCAGCCTCCCTTTCTCTGTCAGTGGGGACCTCATCAGCCAAGCTGGAAGCCATTAATGAACTCATTCGTT
 161 TTGACCATGTATACACCAAGCCTCTAGTCTTAGAGATCCCCTCTGAGACAGAGAGCCAAACTAATGTGGTAGTGAAAATT
 241 GAGGAAGCACCTCTAAGCTCTTCAGAAGAGGATCACCCTGAATTCATTGTCTCAGTGAAGAAAGAACCTTTGGATGATGA
 321 CTTCATTCCCGAGCTGGGCATCTCAAACCTGCTTTCATCCAGCCATTGTCTGAGACCACCTTCCTGCCTGCTGGATGCTC
 401 ACAGTGACTGTGGATATGAGGGCTCCCCTTCTCCCTTCAGCGACATGTCTTCTCCACTTGGTACAGACCACTCCTGGGAG
```

```
 481 GACACTTTTGCCAACGAACTTTTCCCCCAGCTGATTAGTGTCTAAAGCCACCCACCACTGGGCTCCTTCCCTGATCATCA
 561 CACTGCCTAGAGGATAGCATAGGCCTGTCTGCTTCACTAAAAGCCAAAGTAGAGGCTATCTGGCCTTATAAGAATTCCTC
 641 TAAAGTATTTCAAACCTCTTAGATGACTTCCAAGTATTGTCTTTTGACACTCAGCTGTCTGAGGTCTTCAAAGGTATTCC
 721 AATACTACAGCTTTTGAGATTCTCATTATCTTAAAGGTGGTAGCATGCTCTAAATCATAGGGAAAGTCATCTGACAGTTA
 801 TCGTTCAGCCTGGCTATGTAGCCGAGGCTAAGCTGAAACTTGTGACCCTCTTGACCCCACTCCCAAGTGCTGGACTTTAC
 881 CAGGTGTGCAGCTCCACACCGGCCTCTTCACATGTCCTGAAGTAGACATGAGAGTCACCAGTTCTTTCTCTCCTCCCCGC
 961 CCCACAGGTTTCTTTTGTTTCCTTCTACAAGCAGAGAAACAGCAACCTGAGGGGCCTGTCCTTCCTTATGTCCAGTTCAA
1041 GTGAAGATCAAGAATCTTTGTAAAATTATTGGAAATTTACTGTG (SEQ ID NO: 17).
```

Its expression is increased 2-fold in PPARγL treated rats.

HEPATO4

HEPATO4 is a novel 514 bp gene fragment.

The nucleic acid was initially identified in a cloned fragment having the following sequence:

```
 1 gtcgactgtg tcnaggacaa gaaccaagac ctgtgtcagc aggagtccgt gaaggcctac
61 cccaccttcc actattacca ctatgggaag ctt (SEQ ID NO: 18).
```

The cloned sequence was assembled into a contig that includes:

```
  1 TTTTTTTTTTTTTTTTTAAATCGTTTAATTTTATTTCAAAATTGTACAAAAGGCCATAGGTGGCTATAAAAAACTTTGTC
 81 TTTAGCACACATGGAAATTCAGAAAGAACAAAATAGGTATCATTCACAGTGTAAAGGAAGGGCTCTCTTGTAGGAGCGGG
161 ATTACAGGTCCTCCCTCCTTTTTTCTAGTCTCTTGAGGTCTCCCTCCCGGAGGGTTCGGATGAAACTGGTAAATCCCAAT
241 TCCGTGCGGTCACTCTCATACTTTTCTACAAGCTTCCCATAGTGGTAATAGTGGAAGGTGGGGTAGGCCTTCACGGACTC
321 CTGCTGACACAGGTCTTGGTTCTTGTCCTTGACACAGTCGACAGCAGCACAAGCAATCTTCCGGTCATCTTTGAAGGCGT
401 CGGCTGTGGCGGTGAAGTGAGGGATGACCTTCTTACAGTGTGGGCACCAAGGGGCATAGAACATGACCAAGGTGTGTTTC
481 TTCTTCTTCAGGGTCTC (SEQ ID NO: 19);

  1 CAAAATTGTACAAAAGGCCATAGGTGGCTATAAAAAACTTTGTCTTTAGCACACATGGAAATTCAGAAAGAACAAAATAG
 81 GTATCATTCACAGTGTAAAGGAAGGGCTCTCTTGTAGGAGCGGGATTACAGGTCCTCCCTCCTTTTTTCTAGTCTCTTGA
161 GGTCTCCCTCCCGGAGGGTTCGGATGAAACTGGTAAATCCCAATTCCGTGCGGTCACTCTCATACTTTTCTACAAGCTTC
241 CCATAGTGGTAATAGTGGAAGGTGGGGTAGGCCTTCACGGACTCCTGCTGACACAGGTCTTGGTTCTTGTCCTTGACACA
321 GTCGACAGCAGCACAGGCAATCTGAAGGTAACAGCTCACACCACTTTCTGGAACTCTGGCGAGTCCTCTCCAAGAGCCCA
401 CATTCTGCCAAGGTCTGCGTGTTAGCCATGCCACTTACAC (SEQ ID NO: 20);
```

and

```
  1 CACTGAATAAATCGTTTAATTTTATTTCAAAATTGTACAAAAGGCCATAGGTGGCTATAAAAAACTTTGTCTTTAGCACA
 81 CATGGAAATTCAGAAAGAACAAAATAGGTATCATTCACAGTGTAAAGGAAGGGCTCTCTTGTAGGAGCGGGATTACAGGT
161 CCTCCCTCCTTTTTTTCTAGTCTCTTGAGGTCTCCCTCCCGGAGGGTTCGGATGAAACTGGTAAATCCCAATTCCGTGCGG
241 TCACTCTCATACTTTTCTACAAGCTTCCCATAGTGGTAATAGTGGAAGGTGGGGTAGGCCTTCACGGACTCCTGCTGACA
321 CAGGTCTTGGTTCTTGTCCTTGACACAGTCGACAGCAGCACAGGCAATCTTACTCCCCCCCCTCGTGCCG (SEQ ID NO: 21).
```

The resulting consensus sequence is:

```
  1 TTTTTTTTTTTTTTTCACTGAATAAATCGTTTAATTTTATTTCAAAATTGTACAAAAGGCCATAGGTGGCTATAAAAAAC
 81 TTTGTCTTTAGCACACATGGAAATTCAGAAAGAACAAAATAGGTATCATTCACAGTGTAAAGGAAGGGCTCTCTTGTAGG
161 AGCGGGATTACAGGTCCTCCCTCCTTTTTTCTAGTCTCTTGAGGTCTCCCTCCCGGAGGGTTCGGATGAAACTGGTAAAT
241 CCCAATTCCGTGCGGTCACTCTCATACTTTTCTACAAGCTTCCCATAGTGGTAATAGTGGAAGGTGGGGTAGGCCTTCAC
321 GGACTCCTGCTGACACAGGTCTTGGTTCTTGTCCTTGACACAGTCGACAGCAGCACAGGCAATCTTCCGGTCATCTCTGA
401 CAGGCGTCAGCTGTGGCGGTGAACTCAGGCGATGACCTTCTCACAATGAGGGCACCAAGGGGCATAGAACATGACCAAGG
481 TCTGCGTGTTAGCCATGCCACTTACACGGGTCTC (SEQ ID NO: 22).
```

Its expression is increased 2-fold in PPARγL treated rats.

HEPATO5

HEPATO5 is a novel 165 bp gene fragment.

The gene fragment is present in the following sequence:

```
  1 tgatcatgga gtgttttgga atccttatga gacaactact aaaatgatgt gcttcaggtt
 61 tcnacaaact catctacaga cagattctaa gagaaaactg gcatttttgt aaagcaacat
121 aggacattct ntttttgcca agagatggac agccgaagtt cctagg (SEQ ID NO: 23).
```

Its expression is increased 2.5-fold in PPARγL treated rats.

HEPATO6

HEPATO6 is a novel 136 bp gene fragment. The nucleic acid was initially identified in a cloned fragment having the following sequence:

```
 1 agatctcctt cgtgactttg ggacccattt catcaccgaa gccgtgcttg ggggcatcta
61 cgagtacacg cttatcatga (SEQ ID NO: 24).
```

The cloned sequence was assembled into a contig that includes:

```
tcatgataagcgtgtactcgtagatgcccccaagcacggcttcggtgatgaaatgggtcc
caaagtcacgaaggagatctcggtattctccatagctgtactccaggggcaggctcttga
ccctctgaaggaattc (SEQ ID NO: 25).
```

The resulting consensus sequence is:

```
  1 TCATGATAAGCGTGTACTCGTAGATGCCCCCAAGCACGGCTTCGGTGATGAAATGGGTCCCAAAGTCACGAAGGAGATCT
 81 CGGTATTCTCCATAGCTGTACTCCAGGGGCAGGCTCTTGACCCTCTGAAGGAATTC (SEQ ID NO: 26).
```

Its expression is increased 1.5-fold in PPARγL treated rats.

HEPATO7

HEPATO7 is a novel 789 bp gene fragment. The nucleic acid was intially identified in a cloned fragment having the following sequence:

```
  1 ggtacccgca cactgcatct ctttttcata gtagcgaccc gggtaaaacc tcgagcagtt
 61 gcctattagg tcgacttcgc cccacctgag tgagtagact ttttggttat ctttttctcg
121 accccatcca gaaatgatgc atctgtcatt cggttggaat agatatggag accatgggac
181 acaggcaggg acagaattga tgagctcaca ttctttcttg cccgggtgtt ttttcatttc
241 aaccaaagct atgtcattct ggtaggtggc tccgttatac ttttcatga (SEQ ID NO: 27).
```

The cloned sequence was assembled into a contig that includes:

```
  1  aactggaata tttattgggg attatatgac agaattctaa aatggcataa actttgaaga
 61  aactttagga actagccagt aagacttgct tcttgccagc atcgaggaga actaattata
121  cagtttcatt tcaattaaag tataaccctg gaaagaagaa gtgcagaaag aaggaggtcg
181  tagcttcaga cattgtattg agaaacaagg ggtcttccca cgtagtagct aatccaatca
241  aaatagctgg ccactctggt gtaaacacct gggaactctg gtttcccaca gttttctccc
301  cagctcacaa tgccccaaac ataagtgaca ttgttgacat ccttgcagac caaggggcct
361  ccagagtctc ctttgcaggc atcaatggac ccatcactgg tacccgcaca ctgcatctct
421  ttttcatagt agcgacccgg gtaaaacctc gagcagttgc ctattaggtc gacttcgccc
481  cacctgagtg agtagacttt ttggttatct ttttctcgac cccatccaga aatgatgcat
541  ctgtcattcg gttggaatag atatggagac catgggacac aggcagggac agaattgatg
601  agctcacatt ctttcttgcc cgggtgtttt ttcatttcaa ccaaagctat gtc
(SEQ ID NO: 28);

  1  aactggaata tttattgggg attatatgac agaattctaa aatggcataa actttgaaga
 61  aactttagga actagccagt aagacttgct tcttgccagc atcgaggaga actaattata
121  cagtttcatt tcaattaaag tataaccctg gaaagaagaa gtgcagaaag aaggaggtcg
181  tagcttcaga cattgtattg agaaacaagg ggtcttccca cgtagtagct aatccaatca
241  aaatagctgg ccactctggt gtaaacacct gggaactctg gtttcccaca gttttctccc
301  cagctcacaa tgccccaaac ataagtgaca ttgttgacat ccttgcagac caaggggcct
361  ccagagtctc ctttgcaggc atcaatggac ccatcactgg tacccgcaca ctgcatctct
421  ttttcatagt agcgacccgg gtaaaacctc gagcagttgc ctattaggtc gacttcgccc
481  cacctgagtg agtagacttt ttggttatct ttttctcgac cccatccaga aatgatgcat
541  ctgtcattcg gttggaatag atatggagac catgggacac aggcanggac agaattg
(SEQ ID NO: 29);

  1  aactggaata tttattgggg attatatgac agaattctaa aatggcataa actttgaaga
 61  aactttagga actagccagt aagacttgct tcttgccagc atcgaggaga actaattata
121  cagtttcatt tcaattaaag tataaccctg gaaagaagaa gtgcagaaag aaggaggtcg
181  tagcttcaga cattgtattg agaaacaagg ggtcttccca cgtagtagct aatccaatca
241  aaatagctgg ccactctggt gtaaacacct gggaactctg gtttcccaca gttttctccc
301  cagctcacaa tgccccaaac ataagtgaca ttgttgacat ccttgcanga caaagggcct
361  ccagagtctc ctttgcaggc atcaatggac ccatcactgg tacccgcaca ctgcatctct
421  ttttcatagt ancgacccgg gtaaaactcg agcagttgcc tattaggtcg acttcgcccc
481  acctgagtga gta (SEQ ID NO: 30);

  1  gcttaactgg aatatttatt ggggattata tgacagaatt ctaaaatggc ataaactttg
```

```
 61  aagaaacttt aggaactagc cagtaagact tgcttcttgc cagcatcgag gagaactaat
121  tatacagttt catttcaatt aaagtataac cctggaaaga agaagtgcag aaagaaggag
181  gtcgtagctt cagacattgt attgagaaac aaggggtctt cccacgtagt agctaatcca
241  atcaaaatag ctggccactc tggtgtaaac acctgggaac tctggtttcc cacagttttc
301  tccccagctc acaatgcccc aaacataagt gacattgttg acatccttgc agaccaaggg
361  gcctccagag tctcctttgc aggcatcaat ggacccatca ctggtacccg cacactgcat
421  ctctttttca tagtagcgac ccgggtaaaa cctcgagcag ttgcctatta agtcgacttc
481  gccccacctg a (SEQ ID NO: 31);

  1  gtgcttaact ggaatattta ttggggatta tatgacagaa ttctaaaatg gcataaactt
 61  tgaagaaact ttaggaacta gccagtaaga cttgcttctt gccagcatcg aggagaacta
121  attatacagt ttcatttcaa ttaaagtata accctggaaa gaagaagtgc agaaagaagg
181  aggtcgtagc ttcagacatt gtattgagaa acaaggggtc ttcccacgta gtagctaatc
241  caatcaaaat agctggccac tctggtgtaa acacctggga actctggttt cccacagttt
301  tctccccagc tcacaatgcc ccaaacataa gtgacattgt tgacatcctt gcagaccaag
361  gggcctccag agtctccttt gcaggcatca atggacccat cactggtacc cgcacactgc
421  atctcttttt catagtaacg acccgggtaa aacctc (SEQ ID NO: 32);

ngtatggacgtctttattagactggctaaggcctaactctcagtgggnagttcagggagt
gagcagagttgtcgttcatgaaaagtataacggagccacctaccagaatgacatagcttt
ggttgaaatgaaaaaacacccgggcaagaaagaatgtgagctcatcaattctgtccctgc
ctgtgtcccatgg (SEQ ID NO: 33);
```

and

```
ccatgggacacaggcagggacagaattgatgagctcacattctttctngcccgggtgttt
tttcatttcaaccaaagctatgtcattctggtaggtggctccgttatacttttcatgaac
gacaactctgctcactccctgaactgccaactgagagttaggctttagcagtctaataaa
gacgtccatacttggtagttgcgatatctacn (SEQ ID NO: 34).
```

The resulting consensus sequence is:

```
  1 GTGCTTAACTGGAATATTTATTGGGGATTATATGACAGAATTCTAAAATGGCATAAACTTTGAAGAAACTTTAGGAACTA
 81 GCCAGTAAGACTTGCTTCTTGCCAGCATCGAGGAGAACTAATTATACAGTTTCATTTCAATTAAAGTATAACCCTGGAAA
161 GAAGAAGTGCAGAAAGAAGGAGGTCGTAGCTTCAGACATTGTATTGAGAAACAAGGGGTCTTCCCACGTAGTAGCTAATC
241 CAATCAAAATAGCTGGCCACTCTGGTGTAAACACCTGGGAACTCTGGTTTCCCACAGTTTTCTCCCCAGCTCACAATGCC
321 CCAAACATAAGTGACATTGTTGACATCCTTGCAGACCAAGGGGCCTCCAGAGTCTCCTTTGCAGGCATCAATGGACCCAT
401 CACTGGTACCCGCACACTGCATCTCTTTTTCATAGTAGCGACCCGGGTAAAACCTCGAGCAGTTGCCTATTAGGTCGACT
481 TCGCCCCACCTGAGTGAGTAGACTTTTTGGTTATCTTTTTCTCGACCCCATCCAGAAATGATGCATCTGTCATTCGGTTG
561 GAATAGATATGGAGACCATGGGACACAGGCAGGGACAGAATTGATGAGCTCACATTCTTTCTTGCCCGGGTGTTTTTTCA
641 TTTCAACCAAAGCTATGTCATTCTGGTAGGTGGCTCCGTTATACTTTTCATGAACGACAACTCTGCTCACTCCCTGAACT
721 GCCAACTGAGAGTTAGGCCTTAGCCAGTCTAATAAAGACGTCCATACNTTGGTAGTTGCGATATCTACN (SEQ ID NO: 35).
```

Its expression is increased 3-fold in PPARγL treated rats.

HEPATO8

HEPATO8 is a novel 141 bp gene fragment. The nucleic acid was initially identified in a cloned fragment having the following sequence:

```
 1 tgatcatgan atccttgtca ccatccttgg gtggcaggaa gaagactctg tcagtgtagg
61 tcttatagtc cagcgtgggg ntgcccgcct cgtgcac (SEQ ID NO: 36).
```

The cloned sequence was assembled into a contig that includes:

```
tcatgacatccttgtcaccatccttggatggcaggaagaagactctgtcagtgtaggtct
tatagtccagcgtggggatgcccgcctcgtgcacatcatttgtctggtcctccatctcga
tcatcaggtctgtgaatn (SEQ ID NO: 37).
```

The resulting consensus sequence is:

```
 1 TGATCATGACATCCTTGTCACCATCCTTGGATGGCAGGAAGAAGACTCTGTCAGTGTAGGTCTTATAGTCCAGCGTGGGG
81 ATGCCCGCCTCGTGCACATCATTTGTCTGGTCCTCCATCTCGATCATCAGGTCTGTGAATN (SEQ ID NO: 38).
```

Its expression is increased 6-fold in PPARγL treated rats.

HEPATO9

HEPATO9 is novel 270 bp gene fragment. The nucleic acid was initially identified in a cloned fragment having the following sequence:

```
  1 acgcgtgcag ggcgctctgc aggccgggct cgcgagcggg acagctgcag gctgaagcaa
 61 gcgagcgctt cgctgctggc tatatccaat gcatgcatga ggtgcacacg ttcgtgtcca
121 cgtgccaagc catcgatgcc actgtctcag ctgaactcct gaaccacctg ctagagtcca
181 tgccgctgnn cagggtagca gctttcggga tcy (SEQ ID NO: 39).
```

The cloned sequence was assembled into a contig that includes:

```
actagtgggcaccgaggccaagctagagancgccgaggtgctggagctgaccgtgcgacn
cgtgcagggcgctctgcaggccgggctcgcgagcgggacagctgcaggctgaagcaagcg
agcgcttcgctgctggctacatccaatgcacgcatgaggtgcac (SEQ ID NO: 40).
```

The resulting consensus sequence is:

```
  1 NGATCCCGAAAGCTGCTACCCTGNNCAGCGGCATGGACTCTAGCAGGTGGTTCAGGAGTTCAGCTGAGACAGTGGCATCG
 81 ATGGCTTGGCACGTGGACACGAACGTGTGCACCTCATGCATGCATTGGATATAGCCAGCAGCGAAGCGCTCGCTTGCTTC
161 AGCCTGCAGCTGTCCCGCTCGCGAGCCCGGCCTGCAGAGCGCCCTGCACGCGTCGCACGGTCAGCTCCAGCACCTCGGCG
241 NTCTCTAGCTTGGCCTCGGTGCCCACTAGT (SEQ ID NO: 41).
```

Its expression is increased 4-fold in PPARγL treated rats.

HEPATO10

HEPATO10 is a novel gene 189 bp fragment. The cloned fragment is:

```
  1 agatcttaag tgaatccttc tggtttaaac agggtctacc agaaaccatg ttatggagag
 61 tgttgaaaca aatccgactt tgcactctta atcggataag tgacttngtg atgaagcatc
121 attcttcaca gaatgctgtg tgctacaaaa tgctgcccaa catcgggcat gctgtcacta
181 acggctagc (SEQ ID NO: 42).
```

Its expression is increased 2-fold in PPARγL treated rats.

HEPATO11

HEPATO11 is a novel 584 bp gene fragment. The nucleic acid was initially identified in a cloned fragment having the following sequence:

```
  1 gtccttttctt gtgtcaatcc ttacgaatct tgtaatatag taccatttaa attggtaaag
 61 tttcaaaggc aagtttagta ggaatgtaca (SEQ ID NO: 43).
```

The cloned sequence was assembled into a contig that includes

```
   1  ttttttttttt ttttttttacg tcaaattcca tttattgtat agttagagtt tcaatatctt
  61  ttcatttggg aaaccaaaag ataagagaat aaatgtacat tcctactaaa cttgcctttg
 121  aaactttacc aatttaaatg gtactatatt acaagattcg taaggattga cacaagaaag
 181  gactgaagga tgtaagacat ggcccatggc tggcaaaacc ggaaaggcaa tggatatatt
 241  tcagcacttc ctatgtcctc aatcaccttt tagaaaatcc atcataagcc agaatgtaca
 301  tggtagatgc tcctcagaac cacctcaagt gacgccacat aactaccgct taggttgctt
 361  cgaactaggt tcaacctctg tggaacccca agtgcctggt ttgagaaggt ggctaaactt
 421  aatgtaattt atagcaaaaa tatacatcat aattgtacct gcaactttta gagacaaaag
 481  tgattaacct ggcactgaca tccctctatc aaatgccggt taattganaa attagaaaat
 541  atcacagcaa tataa (SEQ ID NO: 44);

   1  gggaaaccaa aagataagag aataaatgta cattcctact aaacttgcct ttgaaacttt
  61  accaatttaa atggtactat attacaagat tcgtaaggat tgacacaaga aaggactgaa
 121  ggatgtaaga catggcccat ggctggcaaa accggaaagg caatggatat atttcagcac
 181  ttcctatgtc ctcaatcacc ttttagaaaa tccatcataa gccagaatgt acatggtaga
 241  tgctcctcag aaccacctca agtgacgcca ataactacc gcttaggttg cttcgaacta
 301  ggttcaacct ctgtggaacc ccaagtgcct ggcttgagaa agtggctaaa cttaatgtaa
 361  tttatagcaa aaatatacat cataattgta (SEQ ID NO:45);

   1  gggaaaccaa aagataagag aataaatgta cattcctact aaacttgcct ttgaaacttt
  61  accaatttaa atggtactat attacaagat tcgtaaggat tgacacaaga aaggactgaa
 121  ggatgtaaga catggcccat ggctggcaaa accggaaagg caatggatat atttcagcac
 181  ttcctatgtc ctcaatcacc ttttagaaaa tccatcataa gccagaatgt acatggtaga
```

241 tgctcc (SEQ ID NO:46);

and

```
  1 ccatttattg tatagttaga gtttcaatat cttttcattt gggaaaccaa aagataagag
 61 aataaatgta cattcctact aaacttgcct ttgaaacttt accaatttaa atgatactat
121 attacaagat tcgtaaggat tgacacaaga aaggactgaa ggatgtaaga catggcccat
181 ggctggcaaa accggaaagg caatggatat atttcagcac ttcctatgtc ctcaatcacc
241 ttttagaaaa tccatcataa gccagaatgt acatggtaga tgctcctcag aaccacctca
301 agtgacgcca cataactacc gcttaggttg cttcgaacta ggttcaacct ctgtggaacc
361 ccaagtgcct ggtttgagaa ggtggctaaa cttaatgtaa tttatagcaa aaatatacat
421 cataattgta cctgcaactt ttagagacaa aagtgattaa cctggcactg acatccctct
481 atcaaatgcc ggttaattga aaaattagaa aatatcacag caatataaca ggttggggat
541 cttaatagga aaagaac (SEQ ID NO: 47).
```

The resulting consensus sequence is:

```
  1 TTTTTTTTTTTTTTTTTACGTCAAATTCCATTTATTGTATAGTTAGAGTTTCAATATCTTTTCATTTGGGAAACCAAAAG
 81 ATAAGAGAATAAATGTACATTCCTACTAAACTTGCCTTTGAAACTTTACCAATTTAAATGGTACTATATTACAAGATTCG
161 TAAGGATTGACACAAGAAAGGACTGAAGGATGTAAGACATGGCCCATGGCTGGCAAAACCGGAAAGGCAATGGATATATT
241 TCAGCACTTCCTATGTCCTCAATCACCTTTTAGAAAATCCATCATAAGCCAGAATGTACATGGTAGATGCTCCTCAGAAC
321 CACCTCAAGTGACGCCACATAACTACCGCTTAGGTTGCTTCGAACTAGGTTCAACCTCTGTGGAACCCCAAGTGCCTGGT
401 TTGAGAAGGTGGCTAAACTTAATGTAATTTATAGCAAAAATATACATCATAATTGTACCTGCAACTTTTAGAGACAAAAG
481 TGATTAACCTGGCACTGACATCCCTCTATCAAATGCCGGTTAATTGAAAAATTAGAAAATATCACAGCAATATAACAGGT
561 TGGGGATCTTAATAGGAAAAGAAC (SEQ ID NO: 48).
```

Its expression is decreased 100-fold in PPARγL treated rats.

HEPATO12

HEPATO12 is a novel 45 bp gene fragment. The cloned sequence is:

```
1 gaattcanat cctgattgag gaaaaagcca gcaacaatgg gtacc (SEQ ID NO: 49).
```

Its expression is decreased 1.5-fold in PPARγL treated rats.

HEPATO13

HEPTO13 is a novel 63 bp gene fragment. The cloned sequence is:

```
1 gccatgcagt gtctctcaag tctcgaggct caccgacttc ggtgagaggc tgggcctcga
61 gac (SEQ ID NO: 50).
```

Its expression is increased 5-fold in PPARγL treated rats.

HEPATO14

HEPTO14 is a novel 96 bp gene fragment. The cloned sequence is:

```
 1 tgatcacaca ggcatgtgca tgtgtttgtt aaagacctaa agggatggct ataggtgaaa
61 ctttgtcttg ttcagaacgc tgactcttgc attgca (SEQ ID NO: 51).
```

Its expression is increased 20-fold in PPARγL treated rats.

HEPATO15

HEPATO16 is a novel gene 315 bp fragment. The nucleic acid was initially identified in a cloned fragment having the following sequence:

```
1  aagcttttca ggttctggga tagtggtacg gatgggtggt ctagttttca aatatattct
61 ctgagaggca tctttagtaa aaacaaaatc tgtgacttgc ctaaataaac tagt (SEQ ID NO: 52).
```

The cloned sequence was assembled into a contig that includes:

```
agatctgtcttacgtttcacaaattagcaccagaaatccacatntgaactgtgtgtaagt
tgccatagcaaatcagttatttgaagcttttcaggttctgggatagtggtacggatgggt
ggtctggtttcaaatatattctctgagaggcatctttagtaaaaacaaaatctgtgactt
gcctgaatgaactagtatatgcttagtcaagttgacagggtcagatcaaccgtcagaaac
aaatgcattcatccgtcagccagcacttatcaaatacatacagagtagcttccctgtgcc
ctgcacagccctagg (SEQ ID NO: 53);
```

and

```
agatctgtcttacgtttcacaaattagcaccaganacccgcatgtgnntgtgtgtaagtg
gccatagcaaatcagttatttgaagcttttcaggttctggggtagtggtacggatgggtg
gtctggtttcaaatatattctctgagaggcatctttagtaaaaacaaaatctgtgacttg
cctgaatgaactagt (SEQ ID NO: 54).
```

The resulting consensus sequence is:

```
  1 AGATCTGTCTTACGTTTCACAAATTAGCACCAGAAACCCACATGTGAACTGTGTGTAAGTGGCCATAGCAAATCAGTTAT
 81 TTGAAGCTTTTCAGGTTCTGGGATAGTGGTACGGATGGGTGGTCTGGTTTCAAATATATTCTCTGAGAGGCATCTTTAGT
161 AAAAACAAAATCTGTGACTTGCCTGAATGAACTAGTATATGCTTAGTCAAGTTGACAGGGTCAGATCAACCGTCAGAAAC
241 AAATGCATTCATCCGTCAGCCAGCACTTATCAAATACATACAGAGTAGCTTCCCTGTGCCCTGCACAGCCCTAGG (SEQ
    ID NO: 55).
```

Its expression is increased 2-fold in PPARγL treated rats

HEPATO16

HEPATO16 is a novel gene169 bp fragment. The nucleic acid was initially identified in a cloned fragment having the following sequence:

```
  1 cctagggttc acgcctctag gcctcccccg cccgggctgc gcttgccaac ccgtttctat
 61 agcaacgccc ggcggntcgc tgctccggaa ggggcacccg ccgcagctcc ggggtcctct
121 ctgcgaggta ccggctggcc aaagaggcgc cagtggcctc acagggccc (SEQ ID NO: 56).
```

The cloned sequence was assembled into a contig that includes:

```
tccggagcancgagcccgccgggcgttgctatagaaacgggttggcaagcgcagccccgg
gcgggggaggcctagaggcgtgaaccctaagtcctgcatgggattatttcgaaccgtccg
gtaagcccgaaatttccagccacactggcggccgttactagtggatccgagctcggtacc
caagcttgggcgtaatcatggtcatagctgtttcctgtgtgaaaattgttatccgctcac
aattccacacaacatacgagccggaaaggnattnt (SEQ ID NO: 57).
```

The resulting consensus sequence is:

```
  1 GGGCCCTGTGAGGCCACTGGCGCCTCTTTGGCCAGCCGGTACCTCGCAGAGAGGACCCCGGAGCTGCGGCGGGTGCCCCT
 81 TCCGGAGCAGCGAGCCCGCCGGGCGTTGCTATAGAAACGGGTTGGCAAGCGCAGCCCCGGGCGGGGGAGGCCTAGAGGCG
161 TGAACCCTAAGTCCTGCATGGGATTATTTCGAACCGTCCGGTAAGCCCGAAATTTCCAGCCACACTGGCGGCCGTTACTA
241 GTGGATCCGAGCTCGGTACCCAAGCTTGGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAAATTGTTATCCGCTCAC
321 AATTCCACACAACATACGAGCCGGAAAGGNATTNT (SEQ ID NO: 58).
```

Its expression is increased 10-fold in PPARγL treated rats

HEPATO17

HEPATO17 is a novel 391 bp gene fragment. The nucleic acid was initially identified in a cloned fragment having the following sequence:

```
  1 tcatgagaac ggaggtagag gcagcggggc agccgcttcg aaaccggaga ctttgtgcag
 61 ctacctgtgc ccaccatcca gcagctgtat cactgggact gtggcctggc ctgctctagg
121 atggtgcttc ggtacc (SEQ ID NO: 59).
```

The cloned sequence was assembled into a contig that includes:

```
  1  tgagaacgna ggtagaggca gcggggcagc cgctcgagcc ggaggacttt gtgcagctac
 61  ctgtgcccac catccagcag ctgtatcact gggactgtgg cctggcctgc tctaggatgg
121  tgcttcggta cctgggccag ctggacgata gggagtttga aaatgccctg caggagctgc
181  agctgaccag aagcatctgg accattgacc tggcctacct catgcgtcac tttggcgtga
241  gacaccgctt ctntacccag actctgggtg ttgacaaggg ttacaagaac cagtccttct
301  ataggaaaca ctttgacaca gaggagaccc gngtggaacc agttgttttg cacaagccaa
361  gg (SEQ ID NO:60);
```

and

ccatgggcgccctgcgccacccgggtcatgagaacggaggtagaggcagcggggcagccg
ctcgagccggaggactttgtgcagctacctgtgcccaccatccagcagctgtatcactgg
gactgtggcctggcctgctctagatggtgcttcggtn (SEQ ID NO: 61).

The resulting consensus sequence is:

```
  1 CCTTGGCTTGTGCAAAACAACTGGTTCCACNCGGGTCTCCTCTGTGTCAAAGTGTTTCCTATAGAAGGACTGGTTCTTGT
 81 AACCCTTGTCAACACCCAGAGTCTGGGTANAGAAGCGGTGTCTCACGCCAAAGTGACGCATGAGGTAGGCCAGGTCAATG
161 GTCCAGATGCTTCTGGTCAGCTGCAGCTCCTGCAGGGCATTTTCAAACTCCCTATCGTCCAGCTGGCCCAGGTNACCGAA
241 GCACCATCCTAGAGCAGGCCAGGCCACAGTCCCAGTGATACAGCTGCTGGATGGTGGGCACAGGTAGCTGCACAAAGTCC
321 TCCGGCTCGAGCGGCTGCCCCGCTGCCTCTACCTCCGTTCTCATGACCCGGGTGGCGCAGGGCGCCCATGG (SEQ ID
    NO: 62).
```

HEPATO18

HEPATO18 is a novel 45 bp gene fragment. The cloned sequence is:

1 aantttcccn cgtaacttac nacngnaccn acccttagg ggccc (SEQ ID NO: 63).

Its expression is decreased 2-fold in PPARγL treated rats.

HEPATO19

HEPATO19 corresponds to glyceraldehyde-3-phosphate dehydrogenase (rat; M17701, human; AW197159). Its transcription is increased 10-fold in PPARγL treated rats.

Glyceraldehyde-3-phosphate dehydrogenase is a glycolytic enzyme that catalyzes the conversion of glyceraldehyde-3-phosphate to 1,3-bisphosphoglycerate.

HEPATO20

HEPATO20 corresponds to UDP-glucose dehydrogenase (rat; AB013732). Its transcription is increased 3-fold in PPARγL treated rats.

UDP-glucose dehydrogenase converts UDP-glucose to UDP-glucuronate, which is a component of the glycosaminoglycans, hyaluronan, chondroitin sulfate, and heparan sulfate.

HEPATO21

HEPATO21 corresponds to succinyl-CoA synthetase alpha subunit (rat; J03621). Its transcription is decreased 3-fold in PPARγL treated rats.

Succinyl-CoA synthetase alpha subunit is an enzyme in the citric acid cycle that catalyzes the conversion of succinyl-CoA to succinate to yield GTP.

HEPATO22

HEPATO22 corresponds to UDP-glucosuronosyl transferase (rat; Y00156, human; AW151709). Its transcription is increased 1.5-fold in PPARγL treated rats.

HEPATO23

HEPATO23 corresponds to alanine aminotransferase (rat; D10354, human; AI991749). Its transcription is decreased 2-fold in PPARγL treated rats.

This protein catalyzes the transfer of the alanine amino group to α-ketoglutarate. This reaction yields pyruvate, which is then used in the glycolytic metabolic pathway.

HEPATO24

HEPATO24 corresponds to tyrosine aminotransferase (rat; M18340, human; X52520). Its transcription is decreased 2-fold in PPARγL treated rats.

This protein is a hepatocyte-specific marker of glucocorticoid activity. This enzyme has been used in many studies of liver involvement of type II diabetes to denote which components may be induced by glucocorticoids. It catalyzes the addition of NH2 group to hydroxyphenylpyruvate to form tyrosine.

HEPATO25

HEPATO25 corresponds to catechol-O-methyltransferase (rat; M60754, human; M58525). Its transcription is increased 4-fold in PPARγL treated rats.

This enyzme is present in high levels of liver tissue and is responsible for the degradation of norepinephrine and epinephrine.

HEPATO26

HEPATO26 corresponds to vitamin D binding protein (rat; J05148, human; X03178).

Vitamin D binding protein is highly expressed in the liver tissue and is the major carrier of vitamin D and its metabolites in the blood. This protein is also known as Gc globulin, a member of the multigene family that includes alpha-fetoprotein and albumin.

Its transcription is increased 1.5-fold in PPARγL treated rats.

HEPATO27

HEPATO27 corresponds to submaxillary gland alpha-2-mu globulin (rat; J00738). Its transcription is increased 25-fold in PPARγL treated rats.

Expression of submaxillary gland alpha-2-mu globulin in the liver is correlated with androgenization of the male rat. The serum fraction of this protein binds hydrocarbons and filters them for excretion in the kidney, which may be related to the nephrotoxicity of these compounds. Rat specific protein and major component of tubular hyaline.

HEPATO28

HEPATO28 corresponds to phospholipase C-1 (rat; M20636). Its transcription is decreased 2-fold in PPARγL treated rats.

Phospholipase C-1 (PLC) cleaves phosphatidyl inositol into diacylglycerol and 1,4,5 inositol triphosphate. PLC tends to be activated by G-proteins of subtype Go.

HEPATO29

HEPATO29 corresponds to ribosomal protein L18 (rat; M20156, human; L11566). Its transcription is decreased 2-fold in PPARγL treated rats.

Ribosomal protein is overexpressed in human colon cancer.

HEPATO30

HEPATO30 corresponds to endoplasmic reticulum transmembrane protein (rat; Y07783, human; AI480032). Its transcription is decreased 2-fold in PPARγL treated rats.

Endoplasmic reticulum transmembrane protein is differentially expressed in rat intestine (Dri 42).

HEPATO31

HEPATO31 corresponds to cytochrome P450 IVA2 (rat; P20816, human; AA886832). Its transcription is increased 6-fold in PPARγL treated rats.

HEPATO32

HEPATO32 corresponds to cytochrome p450 IVA3 (rat; M33936, human; AA070718). Its transcription is increased 5-fold in PPARγL treated rats.

HEPATO33

HEPATO33 corresponds to tricarboxylate transport protein (rat; L12016, human; AW192848). Its transcription is increased 9-fold in PPARγL treated rats.

HEPATO34

HEPATO34 corresponds to cytochrome C oxidase polypeptide l-mitochondrial (rat; S79304, human; AI216989). Its transcription is increased 40-fold in PPARγL treated rats.

HEPATO35

HEPATO35 corresponds to 744 bp gene fragment. This fragment has 83% homology to murine glycerol -3- phosphate acyltransferase. It has also been shown to be differentially expressed in adipose tissue in response to N (2-benzoyl-L-phenyl) -2-tyrosine.

HEPATO36

HEPATO36 corresponds to the L1 retrotransposon m1vi-rn38 which is found in CYP4A2. Its transcription is decreased 15-fold in PPARγL treated rats.

HEPATO37

HEPATO37 corresponds to the L1 retrotransposon ORF2 which is found in CYP2A2. Its transcription is decreased 10-fold in PPARγL treated rats.

HEPATO38

HEPATO38 corresponds to the L1 RnB7 repetive DNA element which is found in CYP4A2. Its transcription is decreased 5-fold in PPARγL treated rats.

HEPATO39

HEPATO39 corresponds to the L1 retrotransposon m1vi-rn8 which is found in CYP4A2. Its transcription is decreased 8-fold in PPARγL treated rats.

HEPATO40

HEPATO40 corresponds to LINE DNA containing 7 ORF's. Its transcription is decreased 6-fold in PPARγL treated rats.

HEPATO41

HEPATO41 corresponds to long interspersed repetitive DNA element LINE3. Its transcription is decreased 2-fold in PPARγL treated rats.

HEPATO42

HEPATO42 corresponds to 2.4 kb repeat DNA left terminal region. Its transcription is decreased 100-fold in PPARγL treated rats.

HEPATO43

HEPATO43 corresponds to the Satellite I core DNA line element. Its transcription is decreased 20-fold in PPARγL treated rats.

HEPATO44

HEPATO44 corresponds a novel 326 bp line element found in CYP2C13. The nucleic acid wwas initially identified in a cloned fragement having the following sequence:

Cloned Fragment:

```
1 gctagcattg acctatgtat tgggcatgct ctgaatgtgt ctctcaggag agatct (SEQ ID NO: 64)
```

The cloned sequence was assembled into a contig that includes:

```
  1  ggcacgagaa aaagtaaaaa aaaaagaaag aaacatggca tatatacata tatatatata
 61  tatgtgtgtg tatatctcca ccaaattttn ntnagattga gaagctaaaa agtgcatgct
121  ggaagggacc ggatatagat ctctcctgag agacacatcc agaacatgtc caatacagag
181  gtcaatgcta gcagcaaacc attcaactga gaacaggacc ccctttggga gaattagagg
241  aattagagaa aggactgaaa gagctgaagg ggcttgcaat cccataagga acaacagtgg
```

301 ccaaccancc agagcttcca gggac (SEQ ID NO: 65)

and

gctagcattgacctctntattgnacatgctctggatgtgtctctcagnanagatct (SEQ ID NO: 66)

and

ggatcctgtnctcagttcantggtttgctactagcattgacctctgtgttggacatgctt

tggatgtgtctctcaggagagatctatatctgctccctttcagcatgaaattttaagctt(SEQ ID NO: 67).

The resulting consensus sequence is:

1 GGCACGAGAAAAAGTAAAAAAAAAAAGAAAGAAACATGGCATATATACATATATATATATATATGTGTGTGTATATCTCCA

81 CCAAATTTTNNTNAGATTGAGAAGCTAAAAAGTGCATGCTGAAAGGGACCAGATATAGATCTCTCCTGAGAGACACATCC

161 AGAACATGTCCAATACAGAGGTCAATGCTAGCAGCAAACCATTCAACTGAGAACAGGATCCCCCTTTGGGAGAATTAGAG

241 GAATTAGAGAAAGGACTGAAAGAGCTGAAGGGGCTTGCAATCCCATAAGGAACAACAGTGGCCAACCANCCAGAGCTTCC

321 AGGGAC (SEQ ID NO: 68)

Its transcription is decreased 100-fold in PPARγL treated rats.

GENERAL SCREEENING AND DIAGNOSTIC METHODS

Several of the herein disclosed methods relate to comparing the levels of expression of one or more HEPATO nucleic acids in a test and reference cell populations. The sequence information disclosed herein, coupled with nucleic acid detection methods known in the art, allow for detection and comparison of the various HEPATO transcripts. In some embodiments, the HEPATO nucleic acids and polypeptide correspond to nucleic acids or polypeptides which include the various sequences (referenced by SEQ ID NOs) disclosed for each HEPATO nucleic acid sequence.

In its various aspects and embodiments, the invention includes providing a test cell population which includes at least one cell that is capable of expressing one or more of the sequences HEPATO 1-32, 36-43 and 44or any combination of HEPATO sequences thereof. By "capable of expressing" is meant that the gene is present in an intact form in the cell and is expressed under particular conditions. Expression of one, some, or all of the HEPATO sequences is then detected, if present, and, preferably, measured. Using sequence information provided by the database entries for the known sequences, or the sequence information for the newly described sequences, expression of the HEPATO sequences can be detected (if present) and measured using techniques well known to one of ordinary skill in the art. For example,

sequences within the sequence database entries corresponding to HEPATO sequences, or within the sequences disclosed herein, can be used to construct probes for detecting HEPATO RNA sequences in, *e.g.*, northern blot hybridization analyses or methods which specifically, and, preferably, quantitatively amplify specific nucleic acid sequences. As another example, the sequences can be used to construct primers for specifically amplifying the HEPATO sequences in, *e.g.*, amplification-based detection methods such as reverse-transcription based polymerase chain reaction. When alterations in gene expression are associated with gene amplification or deletion, sequence comparisons in test and reference populations can be made by comparing relative amounts of the examined DNA sequences in the test and reference cell populations.

For HEPATO sequences whose polypeptide product is known, expression can be also measured at the protein level, *i.e.*, by measuring the levels of polypeptides encoded by the gene products described herein. Such methods are well known in the art and include, *e.g.*, immunoassays based on antibodies to proteins encoded by the genes.

Expression level of one or more of the HEPATO sequences in the test cell population is then compared to expression levels of the sequences in one or more cells from a reference cell population. Expression of sequences in test and control populations of cells can be compared using any art-recognized method for comparing expression of nucleic acid sequences. For example, expression can be compared using GENECALLING® methods as described in US Patent No. 5,871,697 and in Shimkets et al., Nat. Biotechnol. 17:798-803.

In various embodiments, the expression of 2, 3, 4, 5, 6, 7,8, 9, 10, 15, 20, 25, 28, 30, 35, 40, or all of the sequences represented by HEPATO 1-31, 36-43 and 44 are measured. If desired, expression of these sequences can be measured along with other sequences whose expression is known to be altered according to one of the herein described parameters or conditions.

The reference cell population includes one or more cells capable of expressing the measured HEPATO sequences and for which the compared parameter is known, *e.g.*, PPARγ expression status. By "PPARγ expression status" is meant that is known whether the reference cell has had contact with a PPARγ ligand, *e.g.* N-(2-benzoylphenyl)-L-tyrosine. Whether or not comparison of the gene expression profile in the test cell population to the reference cell population reveals the presence, or degree, of the measured parameter depends on the

composition of the reference cell population. For example, if the reference cell population is composed of cells that have not been treated with a known PPARγ ligand, a similar gene expression level in the test cell population and a reference cell population indicates the test agent is not a PPARγ ligand. Conversely, if the reference cell population is made up of cells that have been treated with a known PPAR γ ligand , a similar gene expression profile between the test cell population and the reference cell population indicates the test agent is a PPARγ ligand.

In various embodiments, a HEPATO sequence in a test cell population is considered comparable in expression level to the expression level of the HEPATO sequence if its expression level varies within a factor of 2.0, 1.5, or 1.0 fold to the level of the HEPATO transcript in the reference cell population. In various embodiments, a HEPATO sequence in a test cell population can be considered altered in levels of expression if its expression level varies from the reference cell population by more than 1.0, 1.5, 2.0 or more fold from the expression level of the corresponding HEPATO sequence in the reference cell population. In some embodiments, the variation in expression of a particular HEPATO sequence corresponds to the change in expression level observed for the HEPATO sequences in the presence ir absence of a PPARγ ligand as shown un Table 1.

If desired, comparison of differentially expressed sequences between a test cell population and a reference cell population can be done with respect to a control nucleic acid whose expression is independent of the parameter or condition being measured. Expression levels of the control nucleic acid in the test and reference nucleic acid can be used to normalize signal levels in the compared populations. Suitable control nucleic acids can readily be determined by one of ordinary skill in the art.

In some embodiments, the test cell population is compared to multiple reference cell populations. Each of the multiple reference populations may differ in the known parameter. Thus, a test cell population may be compared to a first reference cell population known to have been exposed to a PPARγ ligand, as well as a second reference population known have not been exposed to a PPARγ ligand.

The test cell population that is exposed to, *i.e.*, contacted with, the test PPARγ ligand can be any number of cells, *i.e.*, one or more cells, and can be provided *in vitro, in vivo*, or *ex vivo*.

In other embodiments, the test cell population can be divided into two or more subpopulations. The subpopulations can be created by dividing the first population of cells to create as identical a subpopulation as possible. This will be suitable, in, for example, *in vitro* or *ex vivo* screening methods. In some embodiments, various sub populations can be exposed to a control agent, and/or a test agent, multiple test agents, or, *e.g.*, varying dosages of one or multiple test agents administered together, or in various combinations.

Preferably, cells in the reference cell population are derived from a tissue type as similar as possible to test cell, *e.g.*, liver tissue. In some embodiments, the control cell is derived from the same subject as the test cell, *e.g.*, from a region proximal to the region of origin of the test cell. In other embodiments, the reference cell population is derived from a plurality of cells. For example, the reference cell population can be a database of expression patterns from previously tested cells for which one of the herein-described parameters or conditions (*e.g.*, PPARγ status, screening, diagnostic, or therapeutic claims) is known.

The subject is preferably a mammal. The mammal can be, *e.g.*, a human, non-human primate, mouse, rat, dog, cat, horse, or cow.

SCREENING FOR PPARγ LIGANDS

In one aspect, the invention provides a method of identifying PPARγ ligands. The PPARγ ligand can be identified by providing a cell population that includes cells capable of expressing one or more nucleic acid sequences homologous to those listed in Table 1 as HEPATO 1-32, 36-44 and, optionally, 33-35. The sequences need not be identical to sequences including HEPATO 1-32, 36-44 and, optionally, 33-35, as long as the sequence is sufficiently similar that specific hybridization can be detected. Preferably, the cell includes sequences that are identical, or nearly identical to those identifying the HEPATO nucleic acids shown in Table 1.

Expression of the nucleic acid sequences in the test cell population is then compared to the expression of the nucleic acid sequences in a reference cell population, which is a cell

population that has not been exposed to the test agent, or, in some embodiments, a cell population exposed to the test agent. Comparison can be performed on test and reference samples measured concurrently or at temporally distinct times. An example of the latter is the use of compiled expression information, *e.g.*, a sequence database, which assembles information about expression levels of known sequences following administration of various agents. For example, alteration of expression levels following administration of test agent can be compared to the expression changes observed in the nucleic acid sequences following administration of a control agent, such as N-(2-benzoylphenyl)-L-tyrosine.

Finding an alteration in the level of expression of the nucleic acid sequence in the test cell population compared to the expression of the nucleic acid sequence in the reference cell population that has not been exposed to the test agent indicates the test agent is a PPARγ ligand.

The invention also includes a PPARγ ligand identified according to this screening method, and a pharmaceutical composition comprising the PPARγ ligands so identified.

SCREENING ASSAYS FOR IDENTIFYING A CANDIDATE THERAPEUTIC AGENT FOR TREATING OR PREVENTING A PATHOPHYSIOLOGIES ASSOCIATED WITH THE PPARγ MEDIATED PATHWAY

The differentially expressed sequences disclosed herein can also be used to identify candidate therapeutic agents pathophysiologies associated with the PPARγ mediated pathway. The method is based on screening a candidate therapeutic agent to determine if it converts an expression profile of HEPATO 1-32, 36-44 and, optionally, 33-35 sequences characteristic of a PPARγ response.

In the method a cell is exposed to a test agent or a combination of test agents (sequentially or simultaneously) and the expression of one or more HEPATO sequences is measured. The expression of the HEPATO sequences in the test population is compared to expression level of the HEPATO sequences in a reference cell population whose PPARγ status is known. If the reference cell population contains cells that have not been exposed to a PPARγ ligand, alteration of the extent of the nucleic acids in the test cell population as compared to the reference cell population indicates that the test agent is a candidate therapeutic agent.

In some embodiments, the reference cell population includes cells that have been exposed to a test agent. When this cell population is used, an alteration in expression of the

nucleic acid sequences in the presence of the agent from the expression profile of the cell population in the absence of the agent indicates the agent is a candidate therapeutic agent. In other embodiments the test cell population includes cells that have not been exposed to a PPARγ ligand. For this cell population, a similarity in expression of the HEPATO sequences in the test and control cell populations indicates the test agent is not a canidate therapuetic agent, while a difference suggests it is a canidate.

The test agent can be a compound not previously described or can be a previously known compound but which is not known to be a PPARγ ligand

An agent effective in stimulating expression of underexpressed genes, or in suppressing expression of overexpressed genes can be further tested for its ability to prevent the PPARγ mediated pathophysiology, *e.g.* NIDDM, and as a potential therapeutic useful for the treatment of such pathophysiology. Further evaluation of the clinical usefulness of such a compound can be performed using standard methods of evaluating toxicity and clinical effectiveness of anti-diabetic agents.

SELECTING A THERAPEUTIC AGENT FOR TREATING A PATHOPHYSIOLOGY ASSOCIATED WITH THE PPARγ MEDIATED PATHWAY THAT IS APPROPRIATE FOR A PARTICULAR INDIVIDUAL

Differences in the genetic makeup of individuals can result in differences in their relative abilities to metabolize various drugs. An agent that is metabolized in a subject to act as an PPARγ ligand can manifest itself by inducing a change in gene expression pattern in the subject's cells from that characteristic of a pathophysiologic state to a gene expression pattern characteristic of a non-pathophysiologic state. Accordingly, the differentially expressed HEPATO sequences disclosed herein allow for a putative therapeutic or prophylactic agent to be tested in a test cell population from a selected subject in order to determine if the agent is a suitable PPARγ ligand in the subject.

To identify a PPARγ ligand, that is appropriate for a specific subject, a test cell population from the subject is exposed to a therapeutic agent, and the expression of one or more of HEPATO 1-32, 36-44 and, optionally, 33-35 sequences is measured.

In some embodiments, the test cell population contains a hepatocyte. In other embodiments, the agent is first mixed with a cell extract, *e.g.*, an liver cell extract, which contains enzymes that metabolize drugs into an active form. The activated form of the

therapeutic agent can then be mixed with the test cell population and gene expression measured. Preferably, the cell population is contacted *ex vivo* with the agent or activated form of the agent.

Expression of the nucleic acid sequences in the test cell population is then compared to the expression of the nucleic acid sequences a reference cell population. The reference cell population includes at least one cell whose PPARγ status is known. If the reference cell had been exposed to a PPARγ ligand a similar gene expression profile between the test cell population and the reference cell population indicates the agent is suitable for treating the pathophysiology in the subject. A difference in expression between sequences in the test cell population and those in the reference cell population indicates that the agent is not suitable for treating the PPARγ pathophysiology in the subject.

If the reference cell has not been exposed to a PPARγ ligand, a similarity in gene expression patterns between the test cell population and the reference cell population indicates the agent is not suitable for treating the PPARγ pathophysiology in the subject, while a dissimilar gene expression patterns indicate the agent will be suitable for treating the subject.

In some embodiments, a decrease in expression of one or more of the sequences HEPATO: 1, 11, 12, 18, 21, 23, 24, 28, 29, and 35-44 or an increase in expression of one or more of the sequences HEPATO: 2 -10, 13-17, 19, 20, 22, 25, 26, 27, and 30-34 in a test cell population relative to a reference cell population is indicative that the agent is therapeutic.

The test agent can be any compound or composition. In some embodiments the test agents are compounds and composition know to be PPARγ ligands, *e.g.* N-(2-benzoylphenyl)-L-tyrosine

METHODS OF DIAGNOSING PATHOPHYSIOLOGIES ASSOCIATED WITH THE PPARγ MEDIATED PATHWAY

The invention further provides a method of diagnosing a pathophysiology associated with the PPARγ mediated pathway, *e.g.*, non-insulin dependent diabetes mellitus, in a subject. A pathophysiology is diagnosed by examining the expression of one or more HEPATO nucleic acid sequences from a test population of cells from a subject suspected of having the pathophysiology.

Expression of one or more of the HEPATO nucleic acid sequences, *e.g.* HEPATO: 1-32, 36-44 and, optionally, 33-35 is measured in the test cell and compared to the expression of the sequences in the reference cell population. The reference cell population contains at least one cell whose PPAR γ status is known. If the reference cell population contains cells that have not been exposed to a PPARγ ligand, than a similarity in expression between HEPATO sequences in the test population and the reference cell population indicates the subject does not have a PPARγ mediated pathophysiology. A difference in expression between HEPATO sequences in the test population and the reference cell population indicates the reference cell population has a PPAR γ mediated pathophysiology.

Conversely, when the reference cell population contains cells that have been exposed to a PPARγ ligand, a similarity in expression pattern between the test cell population and the reference cell population indicates the test cell population has a PPARγ mediated pathophysiology. A difference in expression between HEPATO sequences in the test population and the reference cell population indicates the subject does not have a PPARγ mediated pathophysiology.

METHODS OF TREATING PATHOPHYSIOLOGIES ASSOCIATED WITH THE PPARγ MEDIATED PATHWAY IN A SUBJECT

Also included in the invention is a method of treating, *i.e,* preventing or delaying the onset of a pathophysiology associated with the PPARγ mediated pathway in a subject, *e.g.*, a human, by administering to the subject an agent which modulates the expression or activity of one or more nucleic acids selected from the group consisting of HEPATO: 1-32, 36-44 and, optionally, 33-35. ""Modulates" is meant to include increase or decrease expression or activity of the HEPATO nucleic acids. Preferably, "modulates" is meant to a alter the expression or activity of the HEPATO nucleic acids in a subject to a level similar to a subject not suffering from the pathophysiology.

The pathophysiologies can be any of the pathophysiologies described herein, *e.g.*, NIDDM. The subject can be,*e.g.* a human, a rodent such as a mouse or rat, a dog or cat.

The herein described HEPATO nucleic acids, polypeptides, antibodies, agonists, and antagonists when used therapeutically are referred to herein as "Therapeutics". Methods of administration of Therapeutics include, but are not limited to, intradermal, intramuscular,

intraperitoneal, intravenous, subcutaneous, intranasal, epidural, and oral routes. The Therapeutics of the present invention may be administered by any convenient route, for example by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (*e.g.*, oral mucosa, rectal and intestinal mucosa, etc.) and may be administered together with other biologically-active agents. Administration can be systemic or local. In addition, it may be advantageous to administer the Therapeutic into the central nervous system by any suitable route, including intraventricular and intrathecal injection. Intraventricular injection may be facilitated by an intraventricular catheter attached to a reservoir (*e.g.*, an Ommaya reservoir). Pulmonary administration may also be employed by use of an inhaler or nebulizer, and formulation with an aerosolizing agent. It may also be desirable to administer the Therapeutic locally to the area in need of treatment; this may be achieved by, for example, and not by way of limitation, local infusion during surgery, topical application, by injection, by means of a catheter, by means of a suppository, or by means of an implant. In a specific embodiment, administration may be by direct injection at the site (or former site) of a malignant tumor or neoplastic or pre-neoplastic tissue.

Various delivery systems are known and can be used to administer a Therapeutic of the present invention including, *e.g.*: (*i*) encapsulation in liposomes, microparticles, microcapsules; (*ii*) recombinant cells capable of expressing the Therapeutic; (*iii*) receptor-mediated endocytosis (*See, e.g.,* Wu and Wu, 1987. *J Biol Chem* 262:4429-4432); (*iv*) construction of a Therapeutic nucleic acid as part of a retroviral or other vector, and the like. In one embodiment of the present invention, the Therapeutic may be delivered in a vesicle, in particular a liposome. In a liposome, the protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithin, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Pat. No. 4,837,028; and U.S. Pat. No. 4,737,323, all of which are incorporated herein by reference. In yet another embodiment, the Therapeutic can be delivered in a controlled release system including, *e.g.*: a delivery pump (*See, e.g.,* Saudek, *et al..,* 1989. *New Engl J Med* 321:574 and a semi-permeable polymeric material (*See, e.g.,* Howard, *et al..,* 1989. *J Neurosurg* 71:105). Additionally, the controlled release system

can be placed in proximity of the therapeutic target (*e.g.*, the brain), thus requiring only a fraction of the systemic dose. *See, e.g.*, Goodson, In: *Medical Applications of Controlled Release* 1984. (CRC Press, Bocca Raton, FL).

In a specific embodiment of the present invention, where the Therapeutic is a nucleic acid encoding a protein, the Therapeutic nucleic acid may be administered *in vivo* to promote expression of its encoded protein, by constructing it as part of an appropriate nucleic acid expression vector and administering it so that it becomes intracellular (*e.g.*, by use of a retroviral vector, by direct injection, by use of microparticle bombardment, by coating with lipids or cell-surface receptors or transfecting agents, or by administering it in linkage to a homeobox-like peptide which is known to enter the nucleus (*See, e.g.*, Joliot, *et al..*, 1991. *Proc Natl Acad Sci USA* 88:1864-1868), and the like. Alternatively, a nucleic acid Therapeutic can be introduced intracellularly and incorporated within host cell DNA for expression, by homologous recombination.

As used herein, the term "therapeutically effective amount" means the total amount of each active component of the pharmaceutical composition or method that is sufficient to show a meaningful patient benefit, *i.e.*, treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, the term refers to that ingredient alone. When applied to a combination, the term refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

The amount of the Therapeutic of the invention which will be effective in the treatment of a particular disorder or condition will depend on the nature of the disorder or condition, and may be determined by standard clinical techniques by those of average skill within the art. In addition, *in vitro* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the overall seriousness of the disease or disorder, and should be decided according to the judgment of the practitioner and each patient's circumstances. Ultimately, the attending physician will decide the amount of protein of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein of the present invention and observe the patient's response. Larger doses of protein of

the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. However, suitable dosage ranges for intravenous administration of the Therapeutics of the present invention are generally about 20-500 micrograms (μg) of active compound per kilogram (Kg) body weight. Suitable dosage ranges for intranasal administration are generally about 0.01 pg/kg body weight to 1 mg/kg body weight. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems. Suppositories generally contain active ingredient in the range of 0.5% to 10% by weight; oral formulations preferably contain 10% to 95% active ingredient.

The duration of intravenous therapy using the pharmaceutical composition of the present invention will vary, depending on the severity of the disease being treated and the condition and potential idiosyncratic response of each individual patient. It is contemplated that the duration of each application of the protein of the present invention will be in the range of 12 to 24 hours of continuous intravenous administration. Ultimately the attending physician will decide on the appropriate duration of intravenous therapy using the pharmaceutical composition of the present invention.

Polynucleotides of the present invention can also be used for gene therapy. Gene therapy refers to therapy that is performed by the administration of a specific nucleic acid to a subject. Delivery of the Therapeutic nucleic acid into a mammalian subject may be either direct (*i.e.*, the patient is directly exposed to the nucleic acid or nucleic acid-containing vector) or indirect (*i.e.*, cells are first transformed with the nucleic acid *in vitro*, then transplanted into the patient). These two approaches are known, respectively, as *in vivo* or *ex vivo* gene therapy. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Any of the methodologies relating to gene therapy available within the art may be used in the practice of the present invention. *See e.g.*, Goldspiel, *et al..*, 1993. *Clin Pharm* 12:488-505.

Cells may also be cultured *ex vivo* in the presence of therapeutic agents or proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes.

ASSESSING EFFICACY OF TREATMENT OF A PATHOPHYSIOLOGIES ASSOCIATED WITH THE PPARγ MEDIATED PATHWAY IN A SUBJECT

The differentially expressed HEPATO sequences identified herein also allow for the course of treatment of a pathophysiology to be monitored. In this method, a test cell population is provided from a subject undergoing treatment for pathophysiologies associated with the PPARγ mediated pathway. If desired, test cell populations can be taken from the subject at various time points before, during, or after treatment. Expression of one or more of the HEPATO sequences, *e.g.*, HEPATO: 1-32, 36-44 and, optionally, 33-35, in the cell population is then measured and compared to a reference cell population which includes cells whose pathophysiologic state is known. Preferably, the reference cells not been exposed to the treatment.

If the reference cell population contains no cells exposed to the treatment, a similarity in expression between HEPATO sequences in the test cell population and the reference cell population indicates that the treatment is efficacious. However, a difference in expression between HEPATO sequences in the test population and this reference cell population indicates the treatment is not efficacious.

By "efficacious" is meant that the treatment leads to a decrease in the pathophysiology in a subject. When treatment is applied prophylactically, "efficacious" means that the treatment retards or prevents a pathophysiology. For example, if the PPARγ mediated pathophysiology is NIDDM, a "efficacious" treatment is one that increases insulin sensitivity in a subject.

Efficaciousness can be determined in association with any known method for treating the particular pathophysiology.

HEPATO NUCLEIC ACIDS

Also provided in the invention are novel nucleic acid comprising a nucleic acid sequence selected from the group consisting of HEPATO: 1-18 and 44, or its complement, as well as vectors and cells including these nucleic acids.

Thus, one aspect of the invention pertains to isolated HEPATO nucleic acid molecules that encode HEPATO proteins or biologically active portions thereof. Also included are

nucleic acid fragments sufficient for use as hybridization probes to identify HEPATO-encoding nucleic acids (*e.g.*, HEPATO mRNA) and fragments for use as polymerase chain reaction (PCR) primers for the amplification or mutation of HEPATO nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

"Probes" refer to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt) or as many as about, *e.g.*, 6,000 nt, depending on use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are usually obtained from a natural or recombinant source, are highly specific and much slower to hybridize than oligomers. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

An "isolated" nucleic acid molecule is one that is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Examples of isolated nucleic acid molecules include, but are not limited to, recombinant DNA molecules contained in a vector, recombinant DNA molecules maintained in a heterologous host cell, partially or substantially purified nucleic acid molecules, and synthetic DNA or RNA molecules. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated HEPATO nucleic acid molecule can contain less than about 50 kb, 25 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of any of HEPATO: 1-18 and 44, or a complement of any of these

nucleotide sequences, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of these nucleic acid sequences as a hybridization probe, HEPATO nucleic acid sequences can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, eds., MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to HEPATO nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having at least about 10 nt and as many as 50 nt, preferably about 15 nt to 30 nt. They may be chemically synthesized and may be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in HEPATO: 1-18 and 44. In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in any of these sequences, or a portion of any of these nucleotide sequences. A nucleic acid molecule that is complementary to the nucleotide sequence shown in HEPATO: 1-18 and 44 is one that is sufficiently complementary to the nucleotide sequence shown, such that it can hydrogen bond with little or no mismatches to the nucleotide sequences shown, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, Von der Waals, hydrophobic interactions, etc. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of HEPATO: 1-18 and 44 *e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically active portion of HEPATO. Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 45%, 50%, 70%, 80%, 95%, 98%, or even 99% identity (with a preferred identity of 80-99%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding

nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below. An exemplary program is the Gap program (Wisconsin Sequence Analysis Package, Version 8 for UNIX, Genetics Computer Group, University Research Park, Madison, WI) using the default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2: 482-489, which in incorporated herein by reference in its entirety).

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of a HEPATO polypeptide. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the present invention, homologous nucleotide sequences include nucleotide sequences encoding for a HEPATO polypeptide of species other than humans, including, but not limited to, mammals, and thus can include, *e.g.*, mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the nucleotide sequence encoding a human HEPATO protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in a HEPATO polypeptide, as well as a polypeptide having a HEPATO activity. A homologous amino acid sequence does not encode the amino acid sequence of a human HEPATO polypeptide.

The nucleotide sequence determined from the cloning of human HEPATO genes allows for the generation of probes and primers designed for use in identifying and/or cloning HEPATO homologues in other cell types, *e.g.*, from other tissues, as well as HEPATO homologues from other mammals. The probe/primer typically comprises a substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence of a nucleic acid

comprising a HEPATO sequence, or an anti-sense strand nucleotide sequence of a nucleic acid comprising a HEPATO sequence, or of a naturally occurring mutant of these sequences.

Probes based on human HEPATO nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a HEPATO protein, such as by measuring a level of a HEPATO-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting HEPATO mRNA levels or determining whether a genomic HEPATO gene has been mutated or deleted.

"A polypeptide having a biologically active portion of HEPATO" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically active portion of HEPATO" can be prepared by isolating a portion of HEPATO: 1-18 and 44, that encodes a polypeptide having a HEPATO biological activity, expressing the encoded portion of HEPATO protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of HEPATO. For example, a nucleic acid fragment encoding a biologically active portion of a HEPATO polypeptide can optionally include an ATP-binding domain. In another embodiment, a nucleic acid fragment encoding a biologically active portion of HEPATO includes one or more regions.

HEPATO VARIANTS

The invention further encompasses nucleic acid molecules that differ from the disclosed or referenced HEPATO nucleotide sequences due to degeneracy of the genetic code. These nucleic acids thus encode the same HEPATO protein as that encoded by nucleotide sequence comprising a HEPATO nucleic acid as shown in, *e.g.*, HEPATO: 1-18 and 44.

In addition to the rat HEPATO nucleotide sequence shown in HEPATO: 1-18 and 44, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of a HEPATO polypeptide may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the HEPATO gene may exist

among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a HEPATO protein, preferably a mammalian HEPATO protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the HEPATO gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in HEPATO that are the result of natural allelic variation and that do not alter the functional activity of HEPATO are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding HEPATO proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence of HEPATO: 1-18 and 44, are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the HEPATO DNAs of the invention can be isolated based on their homology to the human HEPATO nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a soluble human HEPATO DNA can be isolated based on its homology to human membrane-bound HEPATO. Likewise, a membrane-bound human HEPATO DNA can be isolated based on its homology to soluble human HEPATO.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of HEPATO: 1-18 and 44. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250 or 500 nucleotides in length. In another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding HEPATO proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at Tm, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (e.g., 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions is hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C. This hybridization is followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of HEPATO: 1-18 and 44 corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of HEPATO: 1-18 and 44 or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X

SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well known in the art. See, *e.g.*, Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of HEPATO: 1-18 and 44 or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). See, *e.g.*, Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo et al., 1981, *Proc Natl Acad Sci USA 78*: 6789-6792.

CONSERVATIVE MUTATIONS

In addition to naturally-occurring allelic variants of the HEPATO sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced into an HEPATO nucleic acid or directly into an HEPATO polypeptide sequence without altering the functional ability of the HEPATO protein. In some embodiments, the nucleotide sequence of HEPATO: 1-18 and 44 be altered, thereby leading to changes in the amino acid sequence of the encoded HEPATO protein. For example, nucleotide substitutions that result in amino acid substitutions at various "non-essential" amino acid residues can be made in the sequence of HEPATO: 1-18 and 44. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of HEPATO without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the HEPATO proteins of the present invention, are predicted to be particularly unamenable to alteration.

In addition, amino acid residues that are conserved among family members of the HEPATO proteins of the present invention, are also predicted to be particularly unamenable to alteration. As such, these conserved domains are not likely to be amenable to mutation. Other amino acid residues, however, (*e.g.*, those that are not conserved or only semi-conserved among members of the HEPATO proteins) may not be essential for activity and thus are likely to be amenable to alteration.

Another aspect of the invention pertains to nucleic acid molecules encoding HEPATO proteins that contain changes in amino acid residues that are not essential for activity. Such HEPATO proteins differ in amino acid sequence from the amino acid sequences of polypeptides encoded by nucleic acids containing HEPATO: 1-18 and 44, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 45% homologous, more preferably 60%, and still more preferably at least about 70%, 80%, 90%, 95%, 98%, and most preferably at least about 99% homologous to the amino acid sequence of the amino acid sequences of polypeptides encoded by nucleic acids comprising HEPATO: 1-18 and 44.

An isolated nucleic acid molecule encoding a HEPATO protein homologous to can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of a nucleic acid comprising HEPATO: 1-18 and 44, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into a nucleic acid comprising HEPATO: 1-18 and 44 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*,

tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in HEPATO is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a HEPATO coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for HEPATO biological activity to identify mutants that retain activity. Following mutagenesis of the nucleic acid, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

In one embodiment, a mutant HEPATO protein can be assayed for (1) the ability to form protein:protein interactions with other HEPATO proteins, other cell-surface proteins, or biologically active portions thereof, (2) complex formation between a mutant HEPATO protein and a HEPATO ligand; (3) the ability of a mutant HEPATO protein to bind to an intracellular target protein or biologically active portion thereof; (*e.g.*, avidin proteins); (4) the ability to bind ATP; or (5) the ability to specifically bind a HEPATO protein antibody.

In another embodiment, the fragment of the complementary polynucleotide sequence described in claim 1 wherein the fragment of the complementary polynucleotide sequence hybridizes to the first sequence.

In other specific embodiments, the nucleic acid is RNA or DNA. The fragment or the fragment of the complementary polynucleotide sequence described in claim 38, wherein the fragment is between about 10 and about 100 nucleotides in length, *e.g.*, between about 10 and about 90 nucleotides in length, or about 10 and about 75 nucleotides in length, about 10 and about 50 bases in length, about 10 and about 40 bases in length, or about 15 and about 30 bases in length.

ANTISENSE

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of a HEPATO sequence or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about

10, 25, 50, 100, 250 or 500 nucleotides or an entire HEPATO coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a HEPATO protein, or antisense nucleic acids complementary to a nucleic acid comprising a HEPATO nucleic acid sequence are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding HEPATO. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding HEPATO. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding HEPATO disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of HEPATO mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of HEPATO mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of HEPATO mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine,

2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a HEPATO protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The

antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA -DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

RIBOZYMES AND PNA MOIETIES

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave HEPATO mRNA transcripts to thereby inhibit translation of HEPATO mRNA. A ribozyme having specificity for a HEPATO-encoding nucleic acid can be designed based upon the nucleotide sequence of a HEPATO DNA disclosed herein. For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a HEPATO-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, HEPATO mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, HEPATO gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of a HEPATO nucleic acid (*e.g.*, the HEPATO promoter and/or enhancers) to form triple helical structures that prevent transcription of the HEPATO gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of HEPATO can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral

backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of HEPATO can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of HEPATO can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of HEPATO can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of HEPATO can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, Proc. *Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

HEPATO POLYPEPTIDES

One aspect of the invention pertains to isolated HEPATO proteins, and biologically active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-HEPATO antibodies. In one embodiment, native HEPATO proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, HEPATO proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a HEPATO protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the HEPATO protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of HEPATO protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of HEPATO protein having less than about 30% (by dry weight) of non-HEPATO protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-HEPATO protein, still more preferably less than about 10% of non-HEPATO protein, and most preferably less than about 5% non-HEPATO protein. When the HEPATO protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially

free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of HEPATO protein in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of HEPATO protein having less than about 30% (by dry weight) of chemical precursors or non-*HEPATO* chemicals, more preferably less than about 20% chemical precursors or non-HEPATO chemicals, still more preferably less than about 10% chemical precursors or non-HEPATO chemicals, and most preferably less than about 5% chemical precursors or non-HEPATO chemicals.

Biologically active portions of a HEPATO protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the HEPATO protein, *e.g.*, the amino acid sequence encoded by a nucleic acid comprising HEPATO 1-20 that include fewer amino acids than the full length HEPATO proteins, and exhibit at least one activity of a HEPATO protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the HEPATO protein. A biologically active portion of a HEPATO protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

A biologically active portion of a HEPATO protein of the present invention may contain at least one of the above-identified domains conserved between the HEPATO proteins. An alternative biologically active portion of a HEPATO protein may contain at least two of the above-identified domains. Another biologically active portion of a HEPATO protein may contain at least three of the above-identified domains. Yet another biologically active portion of a HEPATO protein of the present invention may contain at least four of the above-identified domains.

Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native HEPATO protein.

In some embodiments, the HEPATO protein is substantially homologous to one of these HEPATO proteins and retains its the functional activity, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail below.

In specific embodiments, the invention includes an isolated polypeptide comprising an amino acid sequence that is 80% or more identical to the sequence of a polypeptide whose expression is modulated in a mammal to which PPARγ ligand is administered.

DETERMINING HOMOLOGY BETWEEN TWO OR MORE SEQUENCES

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. See *Needleman and Wunsch* 1970 *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of a DNA sequence comprising HEPATO: 1-18 and 44.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the

number of matched positions by the total number of positions in the region of comparison *(i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

CHIMERIC AND FUSION PROTEINS

The invention also provides HEPATO chimeric or fusion proteins. As used herein, an HEPATO "chimeric protein" or "fusion protein" comprises an HEPATO polypeptide operatively linked to a non-HEPATO polypeptide. A "HEPATO polypeptide" refers to a polypeptide having an amino acid sequence corresponding to HEPATO, whereas a "non-HEPATO polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the HEPATO protein, *e.g.*, a protein that is different from the HEPATO protein and that is derived from the same or a different organism. Within an HEPATO fusion protein the HEPATO polypeptide can correspond to all or a portion of an HEPATO protein. In one embodiment, an HEPATO fusion protein comprises at least one biologically active portion of an HEPATO protein. In another embodiment, an HEPATO fusion protein comprises at least two biologically active portions of an HEPATO protein. In yet another embodiment, an HEPATO fusion protein comprises at least three biologically active portions of an HEPATO protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the HEPATO polypeptide and the non-HEPATO polypeptide are fused in-frame to each other. The non-HEPATO polypeptide can be fused to the N-terminus or C-terminus of the HEPATO polypeptide.

For example, in one embodiment an HEPATO fusion protein comprises an HEPATO domain operably linked to the extracellular domain of a second protein. Such fusion proteins can be further utilized in screening assays for compounds which modulate HEPATO activity (such assays are described in detail below).

In yet another embodiment, the fusion protein is a GST-HEPATO fusion protein in which the HEPATO sequences are fused to the C-terminus of the GST (*i.e.*, glutathione

S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant HEPATO.

In another embodiment, the fusion protein is an HEPATO protein containing a heterologous signal sequence at its N-terminus. For example, a native HEPATO signal sequence can be removed and replaced with a signal sequence from another protein. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of HEPATO can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is an HEPATO-immunoglobulin fusion protein in which the HEPATO sequences comprising one or more domains are fused to sequences derived from a member of the immunoglobulin protein family. The HEPATO-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a HEPATO ligand and a HEPATO protein on the surface of a cell, to thereby suppress HEPATO-mediated signal transduction *in vivo*. The HEPATO-immunoglobulin fusion proteins can be used to affect the bioavailability of an HEPATO cognate ligand. Inhibition of the HEPATO ligand/HEPATO interaction may be useful therapeutically for both the treatments of proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the HEPATO-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-HEPATO antibodies in a subject, to purify HEPATO ligands, and in screening assays to identify molecules that inhibit the interaction of HEPATO with a HEPATO ligand.

An HEPATO chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a

chimeric gene sequence (see, for example, Ausubel *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). An HEPATO-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the HEPATO protein.

HEPATO AGONISTS AND ANTAGONISTS

The present invention also pertains to variants of the HEPATO proteins that function as either HEPATO agonists (mimetics) or as HEPATO antagonists. Variants of the HEPATO protein can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of the HEPATO protein. An agonist of the HEPATO protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the HEPATO protein. An antagonist of the HEPATO protein can inhibit one or more of the activities of the naturally occurring form of the HEPATO protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the HEPATO protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the HEPATO proteins.

Variants of the HEPATO protein that function as either HEPATO agonists (mimetics) or as HEPATO antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the HEPATO protein for HEPATO protein agonist or antagonist activity. In one embodiment, a variegated library of HEPATO variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of HEPATO variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential HEPATO sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of HEPATO sequences therein. There are a variety of methods which can be used to produce libraries of potential HEPATO variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a

degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential HEPATO sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu Rev Biochem* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucl Acid Res* 11:477.

POLYPEPTIDE LIBRARIES

In addition, libraries of fragments of the HEPATO protein coding sequence can be used to generate a variegated population of HEPATO fragments for screening and subsequent selection of variants of an HEPATO protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a HEPATO coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the HEPATO protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of HEPATO proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify HEPATO variants (Arkin and Yourvan (1992) PNAS 89:7811-7815; Delgrave *et al.* (1993) Protein Engineering 6:327-331).

ANTI-HEPATO ANTIBODIES

An isolated HEPATO protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind HEPATO using standard techniques for polyclonal and monoclonal antibody preparation. The full-length HEPATO protein can be used or, alternatively, the invention provides antigenic peptide fragments of HEPATO for use as immunogens. The antigenic peptide of HEPATO comprises at least 8 amino acid residues of the amino acid sequence encoded by a nucleic acid comprising the nucleic acid sequence shown in HEPATO: 1-18 and 44 and encompasses an epitope of HEPATO such that an antibody raised against the peptide forms a specific immune complex with HEPATO. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of HEPATO that are located on the surface of the protein, *e.g.*, hydrophilic regions. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, Proc. Nat. Acad. Sci. USA 78: 3824-3828; Kyte and Doolittle 1982, J. Mol. Biol. 157: 105-142, each incorporated herein by reference in their entirety.

HEPATO polypeptides or derivatives, fragments, analogs or homologs thereof, may be utilized as immunogens in the generation of antibodies that immunospecifically-bind these protein components. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} and $F_{(ab')2}$ fragments, and an F_{ab} expression library. Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies to an HEPATO protein sequence, or derivatives, fragments, analogs or homologs thereof. Some of these proteins are discussed below.

For the production of polyclonal antibodies, various suitable host animals (*e.g.*, rabbit, goat, mouse or other mammal) may be immunized by injection with the native protein, or a

synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, recombinantly expressed HEPATO protein or a chemically synthesized HEPATO polypeptide. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface active substances (*e.g.*, lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), human adjuvants such as *Bacille Calmette-Guerin* and *Corynebacterium parvum*, or similar immunostimulatory agents. If desired, the antibody molecules directed against HEPATO can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction.

The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of HEPATO. A monoclonal antibody composition thus typically displays a single binding affinity for a particular HEPATO protein with which it immunoreacts. For preparation of monoclonal antibodies directed towards a particular HEPATO protein, or derivatives, fragments, analogs or homologs thereof, any technique that provides for the production of antibody molecules by continuous cell line culture may be utilized. Such techniques include, but are not limited to, the hybridoma technique (see Kohler & Milstein, 1975 *Nature* 256: 495-497); the trioma technique; the human B-cell hybridoma technique (see Kozbor, *et al.*, 1983 *Immunol Today* 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, *et al.*, 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, *et al.*, 1983. *Proc Natl Acad Sci USA* 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus *in vitro* (see Cole, *et al.*, 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to a HEPATO protein (see *e.g.*, U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see *e.g.*, Huse, *et al.*, 1989 *Science* 246: 1275-1281) to allow rapid and effective identification of

monoclonal F_{ab} fragments with the desired specificity for a HEPATO protein or derivatives, fragments, analogs or homologs thereof. Non-human antibodies can be "humanized" by techniques well known in the art. See *e.g.*, U.S. Patent No. 5,225,539. Antibody fragments that contain the idiotypes to a HEPATO protein may be produced by techniques known in the art including, but not limited to: (*i*) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (*ii*) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (*iii*) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (*iv*) F_v fragments.

Additionally, recombinant anti-HEPATO antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT International Application No. PCT/US86/02269; European Patent Application No. 184,187; European Patent Application No. 171,496; European Patent Application No. 173,494; PCT International Publication No. WO 86/01533; U.S. Pat. No. 4,816,567; European Patent Application No. 125,023; Better *et al.*(1988) *Science* 240:1041-1043; Liu *et al.* (1987) *PNAS* 84:3439-3443; Liu *et al.* (1987) *J Immunol.* 139:3521-3526; Sun *et al.* (1987) *PNAS* 84:214-218; Nishimura *et al.* (1987) *Cancer Res* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; Shaw *et al.* (1988) *J Natl Cancer Inst.* 80:1553-1559); Morrison(1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; U.S. Pat. No. 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J Immunol* 141:4053-4060.

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme-linked immunosorbent assay (ELISA) and other immunologically-mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of a HEPATO protein is facilitated by generation of hybridomas that bind to the fragment of a HEPATO protein possessing such a domain. Antibodies that are specific for one or more domains within a HEPATO protein, *e.g.*, domains spanning the above-identified conserved regions of HEPATO

family proteins, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Anti-HEPATO antibodies may be used in methods known within the art relating to the localization and/or quantitation of a HEPATO protein (*e.g.*, for use in measuring levels of the HEPATO protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies for HEPATO proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antibody derived binding domain, are utilized as pharmacologically-active compounds [hereinafter "Therapeutics"].

An anti-HEPATO antibody (*e.g.*, monoclonal antibody) can be used to isolate HEPATO by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-HEPATO antibody can facilitate the purification of natural HEPATO from cells and of recombinantly produced HEPATO expressed in host cells. Moreover, an anti-HEPATO antibody can be used to detect HEPATO protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the HEPATO protein. Anti-HEPATO antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

HEPATO RECOMBINANT EXPRESSION VECTORS AND HOST CELLS

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding HEPATO protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a linear or circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell

and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, HEPATO proteins, mutant forms of HEPATO, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of HEPATO in prokaryotic or eukaryotic cells. For example, HEPATO can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (1) to increase expression of recombinant protein; (2) to increase the solubility of the recombinant protein; and (3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION

TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. See, Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the HEPATO expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, (1987) *EMBO J* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, HEPATO can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith *et al.* (1983) *Mol Cell Biol* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells. See, *e.g.*, Chapters 16 and 17 of Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv Immunol* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to HEPATO mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub *et al.*, "Antisense RNA as a molecular tool for genetic analysis," Reviews--Trends in Genetics, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and

"recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, HEPATO protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding HEPATO or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) an HEPATO protein. Accordingly, the invention further

provides methods for producing HEPATO protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding HEPATO has been introduced) in a suitable medium such that HEPATO protein is produced. In another embodiment, the method further comprises isolating HEPATO from the medium or the host cell.

KITS AND NUCLEIC ACID COLLECTIONS FOR IDENTIFYING HEPATO NUCLEIC ACIDS

In another aspect, the invention provides a kit useful for examining a pathophysiology associated with a PPARγ-mediated pathway. The kit can include nucleic acids that detect two or more HEPATO sequences. In preferred embodiments, the kit includes reagents which detect 3, 4, 5, 6, 8, 10, 12, 15, 20, 25, 30, 35,40 or all of the HEPATO nucleic acid sequences.

The invention also includes an isolated plurality of sequences which can identify one or more HEPATO responsive nucleic acid sequences.

The kit or plurality may include, *e.g.*, sequence homologous to HEPATO nucleic acid sequences, or sequences which can specifically identify one or more HEPATO nucleic acid sequences.

NUCLEOTIDE POLYMORPHISMS ASSOCIATED WITH HEPATO GENES

The invention also includes nucleic acid sequences that include one or more polymorphic HEPATO sequences. Also included are methods of identifying a base occupying a polymorphic in an HEPATO sequence, as well as methods of identifying an individualized therapeutic agent for treating PPARγ associated pathologies based on HEPATO sequence polymorphisms.

The nucleotide polymorphism can be a single nucleotide polymorphism (SNP). A SNP occurs at a polymorphic site occupied by a single nucleotide, which is the site of variation between allelic sequences. The site is usually preceded by and followed by highly conserved sequences of the allele (e.g., sequences that vary in less than 1/100 or 1/1000 members of the populations). A single nucleotide polymorphism usually arises due to substitution of one nucleotide for another at the polymorphic site. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement

of a purine by a pyrimidine or vice versa. Single nucleotide polymorphisms can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele.

Polymorphic sequences according to the present invention can include those shown in Table 2. Table 2 describes nine HEPATO sequences for which polymorphisms have been identified. The first column of the table lists the names assigned to the sequences in which the polymorphisms occur. The second and third columns list the rat and human GenBank Accession numbers for the respective sequences. The forth column lists the position in the sequence in which the polymorphic site has been found. The fifth column lists the base occupying the polymorphic site in the sequence in the database, *i.e.*, the wildtype. The sixth column lists the alternative base at the polymorphic site. The seventh column lists any amino acid change that occurs due to the polymorphism.

The polymorphic sequence can include one or more of the following sequences: (1) a sequence having the nucleotide denoted in Table 2, column 5 at the polymorphic site in the polymorphic sequence, and (2) a sequence having a nucleotide other than the nucleotide denoted in Table 2, column 5. An example of the latter sequence is a polymorphic sequence having the nucleotide denoted in Table 2, column 6 at the polymorphic site in the polymorphic sequence.

For example, a polymorphism according to the invention includes a sequence polymorphism in the glyceraldehyde-3-phosphate dehydrogenase gene having the nucleotide sequence of GenBank Accession No. M33197, in which the thymine at nucleotide 441 is replaced by cytosine. In some embodiments the polymorphic sequence includes a nucleotide sequence of glyceraldehyde-3-phosphate dehydrogenase gene having the GenBank Accession No. M33197, wherein the nucleotide at 441 is any nucleotide other that thymine.

In some embodiments, the polymorphic sequence includes the full length of any one of the nine genes in Table2. In other embodiments, the polymorphic sequence includes a polynucleotide that is between 10 and 100 nucleotides, 10 and 75 nucleotides, 10 and 50 nucleotides, or 10 and 25 nucleotides in length.

Table 2

Confirmed Gene	Rat Acc #	Human Seq Calling Acc #	Base Position of cSNP	Base Before	Base Afterc	Change Amino Acid Change
Glyceraldehyde-3-phosphate dehydrogenase	M17701	M33197	441	T	C	
UDP-glucose dehydrogenase	AB013732	AF061016	35	G	A	
			36	A	T	
Catechol-O-methyl transferase	M60754	M65212	550	G	A	VAL to MET
			264	C	T	
Ribosomal protein L18	M20156	L11566	413	G	A	GLY to ASP
			228	G	A	
Complement C8 beta subunit	N/A	M16973	349	G	A	GLU to LYS
Endoplasmic reticulum transmembrane protein	Y07783	AF017786	1013	C	T	

The invention also provides a method of identifying a base occupying a polymorphic site in a nucleic acid. The method includes determining the nucleotide sequence of a nucleic acid that is obtained from a subject. The nucleotide sequence is compared to a reference sequence. Difference in the nucleotide sequence in the test sequence relative to the reference sequence indicates a polymorphic site in the nucleic acid.

Polymorphisms are detected in a target nucleic acid from an individual, *e.g.*, a mammal, human or rodent (such as mouse or rat) being analyzed. For assay of genomic DNA, virtually any biological sample (other than pure red blood cells) is suitable. For example, convenient tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. For assay of cDNA or mRNA, the tissue sample must be obtained from an organ in which the target nucleic acid is expressed.

The detection of polymorphisms in specific DNA sequences, can be accomplished by a variety of methods including, *e.g.*, restriction-fragment-length-polymorphism detection based on allele-specific restriction-endonuclease cleavage (Kan and Dozy Lancet ii:910-912 (1978)), hybridization with allele-specific oligonucleotide probes (Wallace et al. Nucl. Acids Res. 6:3543-3557 (1978)), including immobilized oligonucleotides (Saiki et al. Proc. Natl. Acad.

SCI. USA, 86:6230-6234 (1969)) or oligonucleotide arrays (Maskos and Southern Nucl. Acids Res 21:2269-2270 (1993)), allele-specific PCR (Newton et al. Nucl Acids Res 17:2503-2516 (1989)), mismatch-repair detection (MRD) (Faham and Cox Genome Res 5:474-482 (1995)), binding of MutS protein (Wagner et al. Nucl Acids Res 23:3944-3948 (1995), denaturing-gradient gel electrophoresis (DGGE) (Fisher and Lerman et al. *Proc. Natl. Acad. Sci. U.S.A. 80:1579-l* 583 (1983)), single-strand-conformation-polymorphism detection (Orita et al. Genomics 5:874-879 (1983)), RNAase cleavage at mismatched base-pairs (Myers et al. Science 230:1242 (1985)), chemical (Cotton et al. Proc. Natl. w Sci. U.S.A, 8Z4397-4401 (1988)) or enzymatic (Youil et al. Proc. Natl. Acad. Sci. U.S.A. 92:87-91 (1995)) cleavage of heteroduplex DNA, methods based on allele specific primer extension (Syvanen et al. Genomics 8:684-692 (1990)), genetic bit analysis (GBA) (Nikiforov et al. &&I Acids 22:4167-4175 (1994)), the oligonucleotide-ligation assay (OLA) (Landegren et al. Science 241:1077 (1988)), the allele-specific ligation chain reaction (LCR) (Barrany Proc. Natl. Acad. Sci. U.S.A. 88:189-1 93 (1991)), gap-LCR (Abravaya et al. Nucl Acids Res 23:675-682 (1995)), radioactive and/or fluorescent DNA sequencing using standard procedures well known in the art, and peptide nucleic acid (PNA) assays (Orum et al., Nucl. Acids Res, 21:5332-5356 (1993); Thiede et al., Nucl. Acids Res. 24:983-984 (1996)).

For the purposes of identifying single nucleotide polymorphisms, "Specific hybridization" or "selective hybridization" refers to the binding, or duplexing, of a nucleic acid molecule only to a second particular nucleotide sequence to which the nucleic acid is complementary, under suitably stringent conditions when that sequence is present in a complex mixture (e.g., total cellular DNA or RNA). "Stringent conditions" are conditions under which a probe will hybridize to its target subsequence, but to no other sequences. Stringent conditions are sequence-dependent and are different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter ones. Generally, stringent conditions are selected such that the temperature is about 5°C lower than the thermal melting point (Tm) for the specific sequence to which hybridization is intended to occur at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the target sequence hybridizes to the complementary probe at equilibrium. Typically, stringent conditions include a salt concentration of at least about 0.01 to about 1.0 M Na ion concentration (or other salts), at pH 7.0 to 8.3. The temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) .

Stringent conditions can also be achieved with the addition of destabilizing agents such as formamide. For example, conditions of 5X SSPE (750 mM NaCl, 50 mM NaPhosphate, 5 mM EDTA, pH 7.4) and a temperature of 25-30°C are suitable for allele-specific probe hybridizations.

"Complementary" or "target" nucleic acid sequences refer to those nucleic acid sequences which selectively hybridize to a nucleic acid probe. Proper annealing conditions depend, for example, upon a probe's length, base composition, and the number of mismatches and their position on the probe, and must often be determined empirically. For discussions of nucleic acid probe design and annealing conditions, see, for example, Sambrook et al., or Current Protocols in Molecular Biology, F. Ausubel *et* al., ed., Greene Publishing and Wiley-Interscience, New York (1987).

Many of the methods described above require amplification of DNA from target samples. This can be accomplished by e.g., PCR. *See generally,* PCR Technology: Principles and Applications for DNA Amplification (ed. H. A. Erlich, Freeman Press, N.Y., N.Y., 1992); PCR Protocols: A Guide to Methods and Applications (eds. Innis, *et al.*, Academic Press, San Diego, Calif., 1990); Mattila *et al.*, Nucleic Acids Res. 19, 4967 (1991); Eckert *et al.*, PCR Methods and Applications 1, 17 (1991); PCR (eds. McPherson *et al.*, IRL Press, Oxford); and U.S. Pat. No. 4,683,202 (each of which is incorporated by reference for all purposes).

Other suitable amplification methods include the ligase chain reaction (LCR), (*See* Wu and Wallace, Genomics 4, 560 (1989), Landegren *et al.*., Science 241, 1077 (1988)), transcription amplification (Kwoh *et al.*, Proc. Natl. Acad. Sci. USA 86, 1173 (1989)), and self-sustained sequence replication (Guatelli *et al.*, Proc. Nat. Acad. Sci. USA, 87, 1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

The invention also provides a method of selecting an individualized therapeutic agent for treating a PPARγ associated pathology, *e.g.*, NIDDM, in a subject using HEPATO polymorphisms. The therapeutic agent can be identified by providing a nucleic acid sample from the subject, determining the nucleotide sequence of at least a portion of one or more of the HEPATO 1-32, 36-44 and, optionally, HEPATO 33-35 sequences, and comparing the

HEPATO nucleotide sequence in the subject to the corresponding HEPATO nucleic acid sequence in a reference nucleic acid sample. The reference nucleic acid sample is obtained from a reference individual (who is preferably as similar to the test subject as possible), whose responsiveness to the agent for treating the PPARγ associated pathology is known. The presence of the same sequence in the test and reference nucleic acid sample indicates the subject will demonstrate the same responsiveness to the agent as the reference individual, while the presence of a different sequence indicates the subject will have a different response to the therapeutic agent.

Similarly, the HEPATO-associated sequence polymorphisms can be used to predict the outcome of treatment for a PPARγ associated pathology, *e.g.*, NIDDM, in a subject. A region of an HEPATO nucleic acid sequence from the subject is compared to the corresponding HEPATO sequence in a reference individual whose outcome in response to the treatment for the PPARγ associated pathology is known. A similarity in the HEPATO sequence in the test subject as compared to the sequence in the reference individual suggests the outcome in the subject will be the same as that of the reference individual. An altered HEPATO sequence in the test and reference individual indicates the outcome of treatment will differ in the subject and reference individuals.

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

CLAIMS

What is claimed is:

1. A method of identifying a ligand for peroxisome proliferator activated receptor gamma (PPARγ), the method comprising;

   (a) providing a test cell population comprising a cell capable of expressing one or more nucleic acid sequences selected from the group consisting of HEPATO: 1-32, 36-43 and 44;

   (b) contacting the test cell population with a test agent;

   (c) measuring expression of one or more of the nucleic acid sequences in the test cell population; and

   (d) comparing the expression of the nucleic acid sequences in the test cell population to the expression of the nucleic acid sequences in a reference cell population comprising at least one cell whose PPARγ expression status is known;

   thereby identifying a ligand for PPARγ.

2. The method of claim 1, wherein the method comprises comparing the expression of five or more of the nucleic acid sequences.

3 The method of claim 1, wherein the method comprises comparing the expression of 20 or more of the nucleic acid sequences.

4. The method of claim 1, wherein the method comprises comparing the expression of 40 or more of the nucleic acid sequences.

5. The method of claim 1, wherein the expression of the nucleic acid sequences in the test cell population is decreased as compared to the reference cell population.

6. The method of claim 1, wherein the expression of the nucleic acid sequences in the test cell population is increased as compared to the reference cell population.

7. The method of claim 1, wherein the test cell population is provided *in vitro.*

8. The method of claim 1, wherein the test cell population is provided *ex vivo* from a mammalian subject.

9. The method of claim 1, wherein the test cell is provided *in vivo* in a mammalian subject.

10. The method of claim 1, wherein the test cell population is derived from a human or rodent subject.

11. The method of claim 1, wherein the test cell includes a hepatocyte.

12. A PPARγ ligand identified according to the method of claim 1.

13. A pharmaceutical composition comprising the PPARγ ligand of claim 12.

14. A method of identifying a candidate therapeutic agent for a pathophysiology associated with the for PPARγ mediated pathway, the method comprising;

(a) providing a test cell population comprising a cell capable of expressing one or more nucleic acid sequences selected from the group consisting of HEPATO: 1-32, 36-43 and 44;

(b) contacting the test cell population with a test agent;

(c) measuring expression of one or more of the nucleic acid sequences in the test cell population; and

(d) comparing the expression of the gene in the test cell population to the expression of the nucleic acid sequences in a reference cell population comprising at least one cell whose PPARγ expression status is known,

thereby identifying a therapeutic agent for a pathophysiology associated with the PPARγ mediated pathway

15. The method of claim 14, wherein the pathophysiology is noninsulin-dependent diabetes mellitus (NIDMM).

16. The method of claim 14, wherein the therapeutic agent is a ligand for the PPARγ.

17. A method of identifying an individualized therapeutic agent suitable for treating a pathophysiology associated with a PPARγ mediated pathway appropriate in a selected subject, the method comprising:

(a) providing from the subject a test cell population comprising cells capable of expressing one or more nucleic acid sequences selected from the group consisting of HEPATO: 1-32, 36-43 and 44;

(b) contacting the test cell population with the therapeutic agent;

(c) measuring expression of one or more of the nucleic acid sequences in the test cell population; and

(d) comparing the expression of the nucleic acid sequences in the test cell population to the expression of the nucleic acid sequences in a reference cell population comprising at least one cell whose PPARγ expression status is known;

thereby identifying a therapeutic agent appropriate for the subject.

18. The method of claim 17, wherein the subject is a human or rodent.

19. A method of diagnosing or determining the susceptibility to a pathophysiology associated with a PPARγ mediated pathway in a subject, the method comprising:

(a) providing from the subject a test cell population comprising cells capable of expressing on or more nucleic acid sequences selected from the group consisting of HEPATO: 1-32, 36-43 and 44;

(b) measuring expression of one or more of the nucleic acid sequences in the test cell population; and

(c) comparing the expression of the nucleic acid sequences in the test cell population to the expression of the nucleic acid sequences in a reference cell population comprising at least one cell from a subject not suffering from pathophysiology associated with the for PPARγ mediated pathway; and

(d) identifying a difference in expression levels of the nucleic acid sequences, if present, in the test cell population and reference cell population,

thereby diagnosing or determining the susceptibility to a pathophysiology associated with the PPARγ mediated pathway in the subject.

20. A method of treating a pathophysiology associated with the PPARγ mediated pathway in a subject, the method comprising administering to the subject an agent that modulates the expression or the activity of one or more nucleic acids selected from the group consisting of HEPATO: 1-32, 36-43 and 44.

21. A method of assessing the efficacy of a treatment of pathophysiology associated with the PPARγ mediated pathway in a subject, the method comprising:

(a) providing from the subject a test cell population comprising cells capable of expressing on or more nucleic acid sequences selected from the group consisting of HEPATO: 1-32, 36-43 and 44;

(b) detecting expression of one or more of the nucleic acid sequences in the test cell population; and

(c) comparing the expression of the nucleic acid sequences in the test cell population to the expression of the nucleic acid sequences in a reference cell population comprising at least one cell from a subject not suffering from the pathophysiology associated with the PPARγ mediated pathway ; and

(d) identifying a diference in expression levels of nucleic acid sequences, if present, in the test cell population and the reference cell population,

thereby assessing the efficacy of treatment of the pathophysiology in the subject.

22. An isolated nucleic acid comprising a nucleic acid sequence selected from the group consisting of a HEPATO 1-18 and 44 nucleic acid sequence, or its complement.

23. A vector comprising the nucleic acid of claim 22.

24. A cell comprising the vector of claim 23.

25. A pharmaceutical composition comprising the nucleic acid of claim 22.

26. A polypeptide encoded by the nucleic acid of claim 22.

27. An antibody which specifically binds to the polypeptide of claim 26

28. A kit which detects two or more of the nucleic acid sequences selected from the group consisting of HEPATOs: 1-44.

29. An array which detects one or more of the nucleic acid selected from the group consisting of HEPATOs: 1-44.

30. A plurality of nucleic acid comprising one or more of the nucleic acid selected from the group consisting of HEPATOs: 1-44.

31. A method of identifying a base occupying a polymorphic site in a nucleic acid, the method comprising:

(a) obtaining a nucleic acid from a subject;

(b) determining at least one portion of a region of nucleotide sequence corresponding to a contiguous region of any one ADIPO nucleotide sequence listed in Table 1;

(c) comparing the determined nucleotide sequence to a reference sequence of the nucleic acid; and

(d) identifying a difference in the determined nucleic acid sequence relative to the reference sequence,

wherein a difference in the determined nucleic acid sequence indicates a polymorphic site in the nucleic acid.

32. The method of claim 31, wherein the subject suffers from or is at risk for, a pathophysiology associated with the PPARγ mediated pathway.

33. The method of claim 31, wherein the presence of the polymorphic site is correlated with the presence of the pathophysiology associated with the PPARγ mediated pathway.

34. The method of claim 31, wherein the nucleic acid is genomic DNA.

35. The method of claim 31, wherein the nucleic acid is cDNA.

36. A nucleic acid sequence 20-100 nucleotides in length comprising the polymorphic site identified in the method of claim 31.

37. The method of claim 31, wherein the nucleic acid is obtained from a plurality of subjects, and a base occupying one of the polymorphic sites is determined in each of the subjects.

38. The method of claim 31, wherein the subject is a human or rodent.

39. A method of identifying an individualized therapeutic agent suitable for treating a PPARγ associated pathology in a subject, the method comprising;

(a) providing a nucleic acid sample from the subject;

(b) determining the nucleotide sequence in said subject nucleic acid sample of at least a portion of one or more nucleic acid sequences selected from the group consisting of HEPATO: 1-32, 36-43 and 44;

(c) comparing the nucleic acid sequence in said subject nucleic acid sample to the corresponding nucleic acid sequence from a reference nucleic acid from a reference individual whose reactivity to said agent is known; and

(d) identifying a difference in the nucleic acid sequence, if present, between the subject sample and the reference nucleic acid sample,

thereby identifying a ligand suitable for the subject.

40. The method of claim 39, wherein the nucleic acid sequence is selected from the group consisting of any one nucleotide sequence listed in Table 2.

41. The method of claim 39, wherein the subject is a human or rodent.

42. A method of determining the efficacy of treatment of a PPARγ associated pathology in a subject, the method comprising:

(a) providing a nucleic acid sample from the subject;

(b) determining the nucleotide sequence of at least a portion of one or more nucleic acid sequences selected from the group consisting of HEPATO: 1-32, 36-43 and 44;

(c) comparing the sequence to the corresponding nucleic acid sequence from a cell population whose responsiveness to said treatment for the PPARγ associated pathology agent is known,

thereby determining the efficacy of treatment of a PPARγ associated pathology in the subject.

43. An isolated polynucleotide selected from the group consisting of:

(a) a nucleotide sequence comprising one or more polymorphic sequences of Table 2;

(b) a fragment of the nucleotide sequence including a polymorphic site in the polymorphic sequence;

(c) a complementary nucleotide sequence comprising a sequence complementary to one or more of the polymorphic sequence of Table 2; and

(d) a fragment of the complementary nucleotide sequence including a polymorphic site in the polymorphic sequence.

44. The polynucleotide of claim 43, wherein the polynucleotide is DNA.

45. The polynucleotide of claim 43, wherein the polynucleotide is RNA.

46. The polynucleotide of claim 43, wherein the polynucleotide is between about 10 and about 100 nucleotides in length.

47. The polynucleotide of claim 43, wherein the polynucleotide is between about 10 and about 75 nucleotides in length.

48. The polynucleotide of claim 43, wherein the polynucleotide is between about 10 and about 50 nucleotides in length.

49. The polynucleotide of claim 43, wherein the polynucleotide is between about 10 and about 25 nucleotides in length.

50. The polynucleotide of claim 43, wherein the polynucleotide is a nucleic acid encoding a polypeptide selected from the group consisting of gluteraldehyde-3-phosphate dehydrogenase, UDP-glucose dehydrogenase, catechol-O-methyl transferease, ribosomal protein L18, complement C8 β subunit and endoplasmic reticulum transmembrane protein.

51. The polynucleotide of claim 43, wherein the polymorphic site includes a nucleotide other than the nucleotide listed in Table 2, column 5 for the polymorphic sequence.

52. The polynucleotide of claim 43, wherein the complement of the polymorphic site includes a nucleotide other than the complement of the nucleotide listed in Table 2, column 5 for the complement of the polymorphic sequence.

53. The polynucleotide of claim 43, wherein the polymorphic site includes the nucleotide listed in Table 2, column 6 for the polymorphic sequence.

54. An isolated allele-specific oligonucleotide that hybridizes to a first polynucleotide at a polymorphic site encompassed therein, wherein the first polynucleotide is chosen from the group consisting of:

(a) a nucleotide sequence comprising one or more polymorphic sequences provided that the polymorphic sequence includes a nucleotide other than the nucleotide recited in Table 2, column 5 for said polymorphic sequence;

(b) a nucleotide sequence that is a fragment of said polymorphic sequence, provided that the fragment includes a polymorphic site in said polymorphic sequence;

(c) a complementary nucleotide sequence comprising a sequence complementary to one or more polymorphic sequences ,provided that the complementary nucleotide sequence includes a nucleotide other than the complement of the nucleotide recited in Table 2, column 5; and

(d) a nucleotide sequence that is a fragment of said complementary sequence, provided that the fragment includes a polymorphic site in the polymorphic sequence.

55. A method of assessing hepatoxicity of a PPARγ ligand in a subject, the method comprising:

(a) providing from the subject a test cell population comprising a cell capable of expressing one or more nucleic acid sequences selected from the group consisting of HEPATO: 1-32, 36-43 and 44;

(b) contacting the test cell population with a test agent;

(c) measuring expression of one or more of the nucleic acid sequences in the test cell population; and

(d) comparing the expression of the nucleic acid sequences in the test cell population to the expression of the nucleic acid sequences in a reference cell population comprising at least one cell whose PPARγ expression status is known;

(e) identifying a diference in expression levels of the nucleic acid sequences, if present, in the test cell population and the reference cell population,

thereby assessing the hepatoxicity of the PPARγ ligand in the subject.